

UNIVERSIDAD AUTÓNOMA DE MADRID  
ESCUELA POLITÉCNICA SUPERIOR



Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación

TRABAJO FIN DE GRADO

CONTROL DE CÁMARAS PTZ PARA  
LA RECONSTRUCCIÓN DE ESCENA  
BASADA EN PUNTOS DE INTERÉS.

Elisa Martín Pérez.  
Tutor: Fulgencio Navarro Fajardo.  
Ponente: Jesús Bescós Cano.

Junio 2016



# CONTROL DE CÁMARAS PTZ PARA LA RECONSTRUCCIÓN DE ESCENA BASADA EN PUNTOS DE INTERÉS.

**Elisa Martín Pérez**

**Tutor: Fulgencio Navarro Fajardo**

**Ponente: Jesús Bescós Cano**



**Video Processing and Understanding Lab**

**Departamento de Ingeniería Informática**

**Escuela Politécnica Superior**

**Universidad Autónoma de Madrid**

**Mes 201x**

Trabajo parcialmente financiado por el Ministerio de Economía y Competitividad del  
Gobierno de España bajo el proyecto TEC2014-53176-R (HAVideo) (2015-2017)







# Resumen

Este trabajo de fin de grado tiene como objetivo final la inicialización de fondo (*Background Initialization*) bajo unas condiciones especiales de cámara móvil y presencia de frente, mediante la creación de una imagen panorámica.

Puesto que se parte de un sistema previamente implementado, la primera tarea a realizar es comprobar su funcionamiento, para así analizar la estrategia a seguir. Por tanto, se ha creado un dataset específico para la evaluación de dicho sistema.

Después se han propuesto una serie de mejoras del sistema. Concretamente, se detecta la presencia de errores, o situaciones susceptibles de dar lugar a errores, y se proponen soluciones para una correcta inicialización del fondo. Por tanto, el nuevo sistema integrará dos nuevos módulos claramente diferenciados: de detección y corrección de errores. Del segundo de ellos se han presentado dos aproximaciones posibles.

Por último, se evalúa el sistema modificado y se compara con el de partida, para comprobar si la mejora propuesta ha resultado tal. Igualmente se comparan las dos mejoras implementadas.

Del proyecto se concluye tanto que era necesaria una mejora del sistema base para la generación del fondo de manera correcta, y que las mejoras propuestas no sólo solventan en gran medida el problema detectado, sino que abren el camino a futuras mejoras en la misma línea.

# Palabras clave

Panorámica, inicialización de fondo, homografía, puntos SURF, cámara PTZ, videovigilancia



# Abstract

The aim of this Final Degree Thesis is background initialization through the creation of a panoramic image, under the circumstances of a mobile camera and foreground presence.

Since the starting point is a previously implemented system, it is necessary to check its performance in order to choose a strategy to follow. Therefore, a specific dataset has been developed to evaluate said system.

Afterwards, several improvements are proposed. Specifically, error presence (or situations in which those could take place) is detected, and some solutions are suggested to achieve a correct background initialization. Thus, the new system contains two differentiated modules: error detection module and error correction module. Two different approaches of this second one are presented.

Lastly, the modified system is evaluated and compared with the initial one, to check whether there is an improvement. The two different improvements are compared against each other as well.

The conclusion of this project is that the initial system needed to be improved in order to achieve a proper background generation. Moreover, not only the proposed improvements solve the original problem, but could serve as a basis for new improvements with the same approach.

# Keywords

Panorama, background initialization, homography, SURF points, PTZ cameras, video surveillance



# Agradecimientos

*En primer lugar, a mi tutor Pencho, que ha hecho que todo parezca mucho más fácil. Muchas gracias por toda tu ayuda y haberme dado esta oportunidad.*

*Este trabajo pone el punto final a un largo camino, que ha resultado bastante complicado, sobre todo hacia el final.*

*No podría haberlo conseguido terminar sin la ayuda de muchos, especialmente de mi familia, que, a pesar de tener que aguantar mis nervios, mis dudas, mis horarios y mis ansiedades me han animado a que siga adelante. Julia, siempre consigues convertir mis problemas en tonterías sin importancia.*

*Durante este tiempo se ha cruzado mucha gente en mi camino. Gracias a los que me han ayudado y puesto las cosas más fáciles, he aprendido mucho. Gracias a los que me han puesto las cosas difíciles, he aprendido más aún.*

*En definitiva, gracias a todos a los que habéis estado las mil veces que he querido rendirme y me habéis dado mil y una razones para no hacerlo.*



# Índice general

<b>Resumen</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Agradecimientos</b>	<b>ix</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos . . . . .	3
1.3. Organización de la memoria . . . . .	3
<b>2. Estado del arte.</b>	<b>5</b>
2.1. Inicialización de modelos de fondo para cámaras estáticas. . . . .	5
2.1.1. Métodos basados en estadísticas temporales (TS). . . . .	5
2.1.2. Métodos basados en subintervalos de intensidad estable. . . . .	6
2.1.3. Métodos basados en completado iterativo del modelo. . . . .	6
2.1.4. Métodos basados en etiquetado óptimo. . . . .	7
2.2. Inicialización de modelos de fondo para cámaras móviles. . . . .	7
2.2.1. Técnicas directas ( <i>pixel-based</i> ) . . . . .	8
2.2.2. Técnicas basadas en características . . . . .	8
2.3. <i>Image Registration</i> . . . . .	9
2.3.1. <i>Frame-to-frame</i> . . . . .	10
2.3.2. <i>Frame-to-global</i> . . . . .	10
<b>3. Evaluación del sistema base.</b>	<b>11</b>
3.1. Marco de evaluación. . . . .	11
3.1.1. Dataset. . . . .	11
3.1.2. Métricas. . . . .	13
3.2. Evaluación y análisis . . . . .	15
3.2.1. Objetivos de la evaluación . . . . .	15
3.2.2. Configuración paramétrica . . . . .	15
3.2.3. Resultados . . . . .	15
3.2.4. Comparativas . . . . .	17
3.3. Conclusiones . . . . .	18

<b>4. Sistema propuesto.</b>	<b>19</b>
4.1. Sistema base. . . . .	19
4.1.1. Inserción . . . . .	20
4.2. Modificaciones introducidas. . . . .	20
4.2.1. Módulo de detección de errores . . . . .	22
4.3. Módulo de corrección de errores . . . . .	25
4.3.1. Corrección por hipótesis de estabilidad . . . . .	26
4.3.2. Corrección por multi-panorámica . . . . .	27
<b>5. Evaluación del sistema propuesto.</b>	<b>29</b>
5.1. Marco de evaluación. . . . .	29
5.1.1. <i>Dataset</i> . . . . .	30
5.1.2. Métricas. . . . .	30
5.2. Evaluación y análisis . . . . .	31
5.2.1. Ajuste de evaluación . . . . .	31
5.2.2. Resultados . . . . .	31
5.2.3. Comparativas . . . . .	35
5.3. Conclusiones . . . . .	37
<b>6. Conclusiones y trabajo futuro.</b>	<b>39</b>
6.1. Conclusiones . . . . .	39
6.2. Trabajo futuro . . . . .	40
<b>Bibliografía</b>	<b>41</b>
<b>A. Dataset evaluación sistema base</b>	<b>47</b>
<b>B. Módulos del sistema base</b>	<b>49</b>
B.0.1. Puntos SURF . . . . .	49
B.0.2. Homografía . . . . .	49
<b>C. Resultados de evaluación</b>	<b>51</b>
<b>D. Panorámicas creadas</b>	<b>57</b>



# Índice de figuras

1.1. Esquema de campos de visión de cámaras fijas (rojo) frente a cámaras PTZ (verde). . . . .	2
3.1. Situaciones complicadas. . . . .	12
3.2. Miniaturas del conjunto de secuencias total. De izquierda a derecha, situaciones: frente en movimiento alejado cámara central, frente en movimiento cercano cámara izquierda, frente estático cámara derecha. . . . .	13
3.3. <i>Ground Truth</i> para la cámara lateral izquierda . . . . .	13
3.4. Comparativa entre la panorámica tomada como <i>ground truth</i> y la panorámica generada a partir de la secuencia <i>CT_FR</i> . . . . .	16
3.5. Panorámica generada para la secuencia <i>LT_ST</i> . . . . .	16
3.6. Muestra de panorámicas generadas para la cámara lateral derecha . . . . .	17
3.7. Comparativa de los resultados de las métricas según la situación reflejada en cada secuencia . . . . .	17
4.1. Esquema del sistema base para generación de panorámicas. . . . .	20
4.2. Diagrama de flujo de sistema base con cajas negras . . . . .	21
4.3. Diagrama de flujo de sistema base con bloques de modificación y sus etapas de influencia . . . . .	21
4.4. Puntos de aplicación del módulo de detección de errores . . . . .	22
4.5. Puntos de aplicación del módulo de corrección de errores. . . . .	25
4.6. Diagrama de estados de la corrección por hipótesis de estabilidad. . . . .	27
4.7. Diagrama de estados de la corrección por hipótesis de multi-panorámica. . . . .	28
5.1. Panorámica creada para la misma secuencia con el sistema base (izquierda) y con una de las soluciones propuestas (derecha). . . . .	30
5.2. Resultado de las métricas para la cámara central . . . . .	32
5.3. Resultado de las métricas para la cámara lateral izquierda . . . . .	32
5.4. Resultado de las métricas para la cámara lateral derecha . . . . .	33
5.5. Resultado de las métricas para la cámara central . . . . .	33
5.6. Resultado de las métricas para la cámara lateral izquierda . . . . .	34
5.7. Resultado de las métricas para la cámara lateral derecha . . . . .	34
5.8. Comparativa de panorámicas generadas para la misma secuencia con los diferentes sistemas: de izquierda a derecha, sistema base, sistema MP y sistema HE . . . . .	36
5.9. Comparativa de panorámicas generadas con los dos sistemas propuestos . . . . .	37

A.1. Miniaturas del conjunto de secuencias total. Por filas, cámaras: central, izquierda y derecha. Por columnas, situaciones: frente en movimiento alejado, frente en movimiento cercano, frente estático. . . . .	48
C.1. Resultado de las métricas para la cámara central . . . . .	56
C.2. Resultado de las métricas para la cámara lateral izquierda . . . . .	56
C.3. Resultado de las métricas para la cámara lateral derecha . . . . .	56
D.1. Panorámica generada a partir de la secuencia $CT\_CL$ , mediante el sistema MP	57
D.2. Panorámica generada a partir de la secuencia $CT\_CL$ , mediante el sistema HE	58
D.3. Panorámica generada a partir de la secuencia $CT\_CL$ , mediante el sistema base	58
D.4. Panorámica generada a partir de la secuencia $LT\_FAR$ , mediante el sistema MP . . . . .	59
D.5. Panorámica generada a partir de la secuencia $LT\_FAR$ , mediante el sistema HE . . . . .	59
D.6. Panorámica generada a partir de la secuencia $LT\_FAR$ , mediante el sistema base . . . . .	60
D.7. Panorámica generada a partir de la secuencia $RT\_CL$ , mediante el sistema MP	60
D.8. Panorámica generada a partir de la secuencia $RT\_CL$ , mediante el sistema HE	61
D.9. Panorámica generada a partir de la secuencia $RT\_CL$ , mediante el sistema base	61

# Índice de tablas

3.1. Tabla resultados cámara central . . . . .	15
3.2. Tabla resultados cámara lateral izquierda . . . . .	16
3.3. Tabla resultados cámara lateral derecha . . . . .	17
5.2. Comparativa individual de cada sistema propuesto con el sistema base . . .	35
5.1. Número de veces que cada sistema ha obtenido la mejor puntuación . . . .	35
5.3. Número de veces que cada sistema ha obtenido la mejor puntuación . . . .	37
C.1. Resultados para la cámara central . . . . .	51
C.2. Resultados para la cámara lateral izquierda . . . . .	52
C.3. Resultados para la cámara lateral izquierda . . . . .	52
C.4. Resultados para la cámara central . . . . .	53
C.5. Resultados para la cámara lateral izquierda . . . . .	53
C.6. Resultados para la cámara lateral derecha . . . . .	54
C.7. Resultados para la cámara central . . . . .	54
C.8. Resultados para la cámara lateral izquierda . . . . .	55
C.9. Resultados para la cámara lateral derecha . . . . .	55



# Capítulo 1

## Introducción

En la actualidad, muchos ámbitos requieren de un sistema de videovigilancia que garantice la seguridad en el entorno. Estos sistemas están compuestos habitualmente a su vez por una serie de sistemas, como son:

- El sistema de captura, compuesto por los sensores que reciben la información del medio. Habitualmente en el área de la video vigilancia estos sensores serán cámaras, aunque su naturaleza puede ser muy diversa: móviles o fijas, en color o blanco y negro, de imagen térmica o de profundidad, etc.
- El sistema de procesamiento, esto es, las máquinas y los algoritmos que se encargan del análisis de las secuencias capturadas.
- El sistema de visualización, donde los responsables podrán observar los resultados arrojados por el sistema de procesamiento e interactuar en caso de que fuese necesario.

### 1.1. Motivación

Los sistemas de captura suelen emplear cámaras estáticas, aunque es cada vez más frecuente el uso de diferentes tipos de sensores. Estas cámaras capturan una vista fija, lo cual facilita la tarea al sistema de procesamiento puesto que las condiciones de captura son más estables y controladas.

Sin embargo, la utilidad de los algoritmos de videovigilancia sobre estos medios de captura viene limitada por el reducido campo de visión que ofrece una cámara estática. Si se requiere cubrir grandes áreas, o seguir objetivos a largo plazo, los sistemas de cámaras estáticas están muy limitados.

Como respuesta a estas necesidades, surgen modificaciones del sistema de captura. Una de estas modificaciones es la utilización de cámaras de tipo PTZ (pan-tilt-zoom). Este tipo de cámaras son capaces de tomar imágenes realizando barridos tanto en horizontal como en vertical, así como de hacer zoom en diferentes zonas de la imagen. Con estos medios de

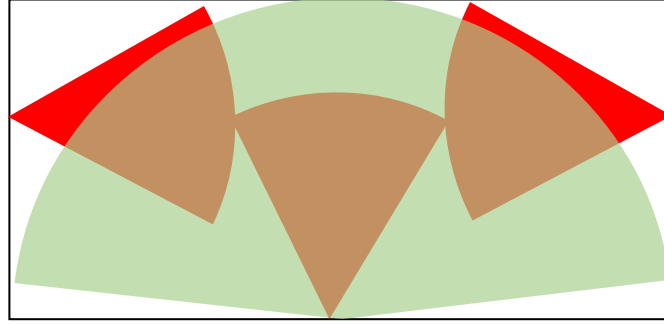


Figura 1.1: Esquema de campos de visión de cámaras fijas (rojo) frente a cámaras PTZ (verde).

captura, será posible paliar algunas de las inconveniencias presentadas por los sistemas de captura mediante cámaras estáticas, ver Figura 1.1.

Sin embargo, la migración a estos sistemas no es directa, y es necesario adaptar los algoritmos del sistema de procesamiento, para que funcionen correctamente con estas nuevas condiciones.

Actualmente ya existen trabajos en este sentido [1]. En dicho trabajo se busca aprovechar las características de las cámaras PTZ para la detección de intrusiones. Para ello, como para numerosas tareas de videovigilancia [2, 3], la primera etapa llevada a cabo es la generación de un modelo de fondo o *Background Initialization* (BGI). En este proyecto se inicializaba un modelo de fondo mediante la generación de una panorámica donde se iban concatenando las vistas que iba cubriendo la cámara móvil.

Las tareas de BGI han sido extensamente estudiadas en el estado del arte. Los principales problemas que presentan estas tareas están relacionados con lo que se denomina “inicio en caliente”. Se entiende por “inicio en caliente” a la generación de un modelo de fondo cuando no es posible visualizar el mismo por completo en el momento de la inicialización. Esta tarea supone un reto aún no resuelto para el estado del arte, si bien para secuencias obtenidas mediante cámaras estáticas los resultados son notables.

El trabajo desarrollado en [1] generaba el mencionado modelo de fondo sin “inicio en caliente”. Por lo tanto, los resultados del modelo generado eran de poca calidad cuando no se conseguía un fondo sin la presencia de ningún elemento de frente. Puesto que [1] se aplicaba sobre cámaras situadas en un entorno de alta concurrencia, la probabilidad de que fuera necesario un inicio en caliente es muy elevada. Adicionalmente, dado que el área cubierta es más amplia que el de una cámara estática, y por tanto el modelo de fondo es mucho mayor, las probabilidades del “inicio caliente” se magnifican.

Será por tanto la motivación de este proyecto los problemas de BGI en secuencias obtenidas mediante cámaras PTZ.

## 1.2. Objetivos

El objetivo de este trabajo de fin de grado es el de intentar solucionar los problemas del BGI dadas las condiciones mencionadas, de cámara móvil e “inicio en caliente”. Para ello, se proponen los siguientes objetivos parciales:

- En primer lugar, se llevarán a cabo una serie de pruebas a modo de verificación de hipótesis –motivación– del proyecto. Con ello se pretende demostrar los defectos del sistema desarrollado y los efectos que tienen sobre los modelos de fondo generados.
- Posteriormente, se pretende diseñar e implementar un módulo para la detección automática de defectos en la creación del modelo de fondo. Para ello se analizarán características propias de la inicialización del fondo como áreas con alto movimiento. Se trabajará con medidas de calidad en la creación de panorámicas, como asociación de puntos o proyección inversa. También se realizará un análisis de la información contextual y de la información aportada por un conocimiento previo del entorno de aplicación.
- A continuación, se procederá a desarrollar un módulo para la corrección de los fallos detectados. Se hará uso tanto de información contextual como de rutas realizadas. Con esto, se tratará de hacer más eficiente la reinserción de imágenes para el modelo de fondo. Puesto que no se podrá generar un modelo completo del fondo hasta que no se haya podido visualizar y captura por completo, el módulo deberá encargarse de aprovechar al máximo las situaciones en las que esto sea posible, para minimizar los costes de generación del fondo de escena.
- Por último, se llevarán a cabo una serie de pruebas que permitan verificar que el trabajo realizado ha satisfecho la motivación inicial del proyecto.

## 1.3. Organización de la memoria

La memoria consta de los siguientes capítulos:

- Capítulo 1. Motivación, objetivos del proyecto y estructura de la memoria.
- Capítulo 2. Estado del arte de técnicas de inicialización de modelos de fondo, tanto para cámaras estáticas como móviles, y de *image registration*.
- Capítulo 3. Evaluación del sistema base. Extracción de conclusiones a partir de los resultados
- Capítulo 4. Descripción detallada del sistema propuesto. Propuesta de dos mejoras diferenciadas.

- Capítulo 5. Evaluación del sistema propuesto. Comparativa con el sistema base. Comparativa entre ambos sistemas propuestos.
- Capítulo 6. Conclusiones finales y sugerencia de posibles líneas de trabajo futuro.
- Bibliografía y Anexos.



## Capítulo 2

# Estado del arte.

Para enfrentar los problemas derivados de la inicialización del modelo de fondo existen distintas aproximaciones. Las más comunes en el estado del arte están enfocadas a tareas sobre cámaras estáticas. Sin embargo, cuando se trata de aplicar estas técnicas de cámaras estáticas sobre cámaras móviles los resultados son muy inferiores. Por ello, se encuentran grandes diferencias entre algoritmos de inicialización de fondo según estén orientados a un tipo u otro de sensores de captura. A continuación se van a detallar algunas de las técnicas estudiadas, tanto para cámaras estáticas (2.1) como para móviles (2.2). Posteriormente, la gestión de nuevos *frames* requerirá de un procesado previo en el caso de cámaras móviles. Éste procesado se conoce como registrado de imagen (*image registering*), y consistirá en ubicar espacialmente dicho *frame* en el fondo inicializado. Algunas técnicas del estado del arte se detallarán en la sección 2.3.

### 2.1. Inicialización de modelos de fondo para cámaras estáticas.

#### 2.1.1. Métodos basados en estadísticas temporales (TS).

Este tipo de métodos parten de la premisa de que, en una secuencia de imágenes, para cada pixel el valor más frecuente será el valor de fondo. Uno de los primeros métodos,[4], utilizaba un simple sistema de votación para asignar a cada píxel el valor más repetido en la secuencia temporal. Posteriormente, se generalizó el uso de la mediana para la asignación de valores a los píxeles [5, 6, 7, 8]. Sin embargo, el coste computacional de estas técnicas era muy elevado. Más adelante destaca la técnica de *Mixture of Gaussians* (MoG), que modela cada pixel mediante una mezcla de Gaussianas [9]. Se trata de una técnica eficiente pero propensa a dar falsos positivos de fondo en objetos del frente. En [10], se elimina la necesidad de parámetros asignando una probabilidad a cada píxel tras haber observado qué conjunto de valores había tomado previamente. Por último, cabe destacar [11], que hace uso de una

red neuronal artificial, cuyo aprendizaje inicial se toma como la estimación inicial del fondo. Entre las desventajas de los métodos que siguen este criterio se encuentran la necesidad de visualizar la totalidad del fondo de forma simultánea y la carencia de coherencia espacial de la reconstrucción.

### 2.1.2. Métodos basados en subintervalos de intensidad estable.

Se parte de la premisa de que el valor del fondo se corresponde con el valor estable durante más tiempo en una secuencia temporal de imágenes. Los algoritmos que siguen esta metodología se desarrollan en dos etapas. En primer lugar se fijan los “intervalos estables” a nivel de píxel. Posteriormente, se decide entre los intervalos estables cuál será el valor de fondo según un determinado criterio. El método de *Adaptive Smoothness Detector* [12] propone una técnica heurística para seleccionar el valor de estabilidad más probable de ser fondo. Este método tiene problemas con secuencias cortas y movimiento constante. Muy similar es lo propuesto en *Local Image Flow* [13], que evalúa cada una de las hipótesis de estabilidad mediante el flujo óptico del vecindario de cada píxel, y se elige como fondo la hipótesis más probable. Pese a ser un buen estimador, los métodos en los que se usa flujo óptico resultan muy costosos computacionalmente y sensibles al ruido. En [14] se presenta un método en el que los intervalos estables para cada píxel se detectan mediante un algoritmo de región creciente, tomando como semillas los mínimos locales de varianza de intensidad en ventanas temporales. En la segunda etapa se realiza una propagación espacial de los clústers utilizados en la primera etapa hasta conseguir estabilidad. El problema de este método es la gran cantidad de tiempo de procesamiento. Al igual que ocurría para los métodos basados en estadísticas temporales este método requiere la visualización de la totalidad del fondo de forma simultánea, además de resultar problemático frente a fondos estáticos. Sin embargo, la metodología basada en la búsqueda de intervalos de intensidad estable resuelve la incoherencia espacial a la que se enfrentaban dichos métodos.

### 2.1.3. Métodos basados en completado iterativo del modelo.

Estos modelos de nuevo siguen un proceso de dos etapas; en la primera se eligen regiones de la imagen que se toman como referencia para el fondo y en la segunda se completa este modelo iterativamente. [15] divide las imágenes en dos, por un lado está el llamado “fondo seguro”, y por otro la denominada “zona de confusión”. En cada iteración, se buscan áreas de fondo en la zona de confusión y se expande la región clasificada como fondo hasta que la zona de confusión esté vacía o resulte imposible continuar. [16] identifica bloques que forman parte del fondo mediante clasificación online para posteriormente integrarlos en un modelo completo. Inicialmente se entrena una colección de bloques de fondo extraídas de diferentes secuencias, y se genera una función de decisión. Mediante el uso de diferencias *inter-frame* y flujo óptico, un proceso clasifica los bloques y verifica que la actualización ha sido llevada

a cabo correctamente. Un último ejemplo de este tipo de métodos es [17], que se ejecuta a nivel de bloque y de forma no recursiva, con el objetivo de conseguir el mejor modelo de fondo a partir del mínimo número posible de *frames*. Para ello, se dividen las imágenes en bloques y se fijan aquellos que forman parte del fondo. A continuación se utilizan bloques cuya frecuencia espectral sea coherente con los bloques vecinos que forman parte del fondo, utilizando la transformada Hadamard, para completar el fondo. En comparación con los métodos anteriormente comentados este método se encuentra en clara ventaja, al resolver todos los problemas mencionados: ya no se encuentran problemas respecto a la coherencia espacial ni al frente estático, y no se requiere visualizar por completo el fondo de forma simultánea. No obstante, la problemática se encuentra en la inicialización de semillas, que al ser un procedimiento crítico se convierte en compleja.

#### 2.1.4. Métodos basados en etiquetado óptimo.

El último método consiste en la selección del valor (etiqueta) óptimo para cada posición de la imagen mediante una minimización de energía. Las técnicas de minimización de energía propuestas en el estado del arte se basan en asunciones sobre características del fondo. Por ejemplo, en [18] se propone una minimización de gradientes. En [19] se busca minimizar tanto la energía de gradientes como espacial y de color. En [20] se busca optimizar una función que incluye términos de movimiento además del suavizado de gradientes. Más recientemente, se ha buscado enfrentar mediante esta estrategia problemas de fondos no homogéneos y fondos con objetos estacionarios. Para los fondos no homogéneos [21] propone una técnica de minimización de energía local en función de su información y la de los máximos locales. Para gestionar los objetos estacionarios, [22] propone una función de minimización que combina información a nivel de píxel, de región y de coherencia espacio temporal de ambas. Para llevar a cabo la minimización de estas funciones se usan dos técnicas principalmente: *Loopy Belief Propagation* [23] y *Graph cuts* [24]. Al igual que los métodos basados en completado iterativo, estos métodos no presentan problemas de coherencia espacial, frente estático o visualización completa de fondo. En este caso la dificultad radica en que las técnicas de minimización de energía están fuertemente relacionadas con la estimación de posición.

## 2.2. Inicialización de modelos de fondo para cámaras móviles.

Dentro del área de las cámaras estáticas existen diversas técnicas para realizar una inicialización adecuada del fondo, o para corregir la más básica que consistiría en capturar un único frame y fijarlo como fondo. Sin embargo, para la inicialización de un modelo de fondo en cámaras móviles es necesario un procesamiento de las imágenes capturadas, que consiste en solapar las partes comunes de las imágenes para construir lo que se conoce como panorámica de fondo. Este proceso se denomina *stitching*, y se aborda en el estado del

arte siguiendo principalmente dos estrategias, basadas en diferencias a nivel de píxel en el solapado 2.2.1, y basadas en correspondencia de características locales o *local features* 2.2.2.

### 2.2.1. Técnicas directas (*pixel-based*)

Las técnicas directas, también denominadas *pixel-based*, comparan la intensidad de cada píxel de la zona de solape de dos imágenes. Si bien esta estrategia es la más simple, su coste computacional es muy alto, ya que hay que comparar uno a uno cada píxel de cada una de las imágenes. Para saber cuál es la correspondencia más óptima entre píxeles de cada imagen, se llevan a cabo medidas de error como las descritas en [25]. Existen medidas *ad-hoc* como pueden ser SAD (*Sum of Absolute Differences*) o SSD (*Sum of Squared Differences*). Estas medidas de error no funcionan correctamente en los casos en que las áreas de solape no presenten exactamente la misma información. Para resolver este problema, en nuevas métricas se asignan pesos a los píxeles, y se calcula la WSSD (*Weighted Sum of Square differences*). Otra situación conflictiva es la variación de la información de la imagen por culpa de la iluminación, véase ajustes automáticos de ganancia de las cámaras, saturaciones, u otros. Estimando la relación entre iluminaciones como lineal, en [26] se propone una regresión lineal para encontrar las similitudes entre imágenes, cuyo coste computacional es algo mayor que la similitud de mínimos cuadrados. Otra manera de encontrar similitudes cuando las intensidades de las imágenes a comparar difieren es mediante la maximización de la correlación entre las dos imágenes (es decir, su correlación cruzada).

Con el objetivo de la mejora del coste computacional en estas técnicas, se introduce el uso de la Transformada de Fourier, altamente eficiente en lo que al coste computacional se refiere. Se puede usar tanto para calcular la SSD como la correlación normalizada.

Como ya se ha comentado la principal desventaja de estas técnicas se encuentra en su alto coste computacional, aunque permiten obtener medidas de la fiabilidad de la reconstrucción que se ha llevado a cabo.

### 2.2.2. Técnicas basadas en características

Una forma computacionalmente menos costosa y que permite estimar de la posición relativa de dos imágenes con un área en común es la técnica basada en características locales. Ésta se compone de dos etapas bien diferenciadas: la obtención de correspondencias locales entre imágenes, y la estimación de la posición relativa de las imágenes en función de dichas correspondencias. Se han desarrollado multitud de técnicas de detección y descripción de características, sin embargo a continuación se van a abordar SIFT y SURF, por ser las más contrastadas del estado del arte para esta tarea.

SIFT [27] lleva a cabo la detección en un proceso de cuatro etapas en el que se crea un espacio-escala híbrido, compuesto de versiones diezmadas de la misma imagen y a la que se le han aplicado sucesivos filtros gaussianos (por lo que se denominan pirámides gaussianas).

Después se buscan puntos de interés mediante DoG (*Difference-of-Gaussians*), que detecta máximos y mínimos de la imagen. Posteriormente se refina la búsqueda desechando puntos que se encuentren en los bordes de la imagen o que estén poco contrastados. Una vez la detección de los puntos de interés se ha llevado a cabo, tiene lugar la descripción, previo paso por la asignación de orientaciones a cada uno de los puntos detectados, mediante el cálculo del gradiente.

SURF [28] nació con el objetivo de mejorar el tiempo de computación de SIFT, buscando para ello nuevas herramientas para la detección y descripción de puntos de interés. El método elegido para la detección es el determinante de la matriz Hessiana, como una aproximación muy básica y aún más simplificada que *DoG* de la Laplaciana. A diferencia del método SIFT, en el que se crea una pirámide con versiones diezmadas y filtradas de la imagen, para mejorar la eficiencia del algoritmo SURF se escoge un nuevo enfoque, mediante el uso de *integral images* y *box filters*. La introducción de estos conceptos permite evitar la aplicación recursiva de un filtro, ya que en este caso se aplicarán sobre la imagen filtros de tamaño variable. Respecto al descriptor de SURF, las orientaciones se obtienen mediante el uso de *Haar wavelets*. Si bien es cierto que SURF consigue lo que se propone, mejorar considerablemente la velocidad de SIFT, su rendimiento en imágenes de bajo contraste y puntos de vista muy diferentes es mejorable [29].

Estas dos técnicas son una herramienta para la obtención de correspondencias entre imágenes. Sin embargo, la estimación de transformaciones en base a esas correspondencias es muy sensible a pequeños errores, por lo que a pesar de su contrastado funcionamiento, habitualmente existen fallos que deben ser corregidos antes de utilizar dicha información. En el estado del arte, por ejemplo en [30], se proponen técnicas basadas en RANSAC (*Random sample consensus*) para eliminar las correspondencias erróneas, y estimar las transformaciones entre las imágenes, homografías, en función de las correspondencias restantes. Éstas encuentran problemas con fondos homogéneos que obtienen un bajo número de correspondencias. También tienen dificultades con grandes áreas de frente como movimiento homogéneo, que dar lugar a correspondencias y estimaciones de transformación erróneas.

Sin embargo se encuentran en clara ventaja en comparación con las técnicas directas ya que su coste computacional disminuye respecto a estas. Al igual que las técnicas directas es posible conocer la fiabilidad de la reconstrucción.

### 2.3. *Image Registration.*

Una vez el proceso de *stitching* está concluido, existen errores que pueden derivarse de esta etapa. Tanto para su corrección como para el uso del modelo de fondo en posteriores aplicaciones, es necesaria una etapa que se conoce como *image registering*. Consiste en hacer corresponder nuevas imágenes con las zonas de la panorámica que se desea comparar, o en caso de querer corregir algún error, sustituir éstas. En el estado del arte, esta tarea se afronta

siguiendo dos estrategias: *frame-to-frame* y *frame-to-global* [31].

### 2.3.1. *Frame-to-frame*

Este método se basa en la reutilización de píxeles mediante su superposición en el *frame* actual y los anteriores. Bajo la premisa de que la superposición siempre es útil en las cámaras PTZ, en [32] se utilizan transformaciones geométricas para alinear *frames* consecutivos cuando la cámara PTZ rota, partiendo de dos alternativas para ello: o bien para cada cambio se calculan todas las matrices correspondientes a todas las posibles combinaciones de pan y tilt o bien se toman como parámetros el tamaño del sensor CCD y la longitud focal (se suele optar por esto último).

Para estimar los parámetros zoom, pan y tilt, el método de [33] utiliza aquellos obtenidos del *frame* anterior, aunque se encuentra con la limitación de acumular error en largas secuencias. En [34] se estiman optimizando el error entre dos *frames* consecutivos, mediante el gradiente. Este método además lleva a cabo una estratificación (pirámides laplacianas), para aportar robustez al algoritmo.

Estas técnicas poseen una imagen panorámica general. Sin embargo, para reubicar cada *frame*, esto es, para llevar a cabo el *image registration*, necesitan conocer información del *frame* anterior. El posicionamiento se realiza con respecto a éste *frame* anterior, por lo que nunca se trabaja como si se tuviese conocimiento del modelo global, sino cuadro a cuadro. Estos sistemas presentan una gran flexibilidad a la hora del movimiento de la cámara, aunque los resultados a nivel del modelo de fondo generado son peores. Esto se debe a que los pequeños errores se tienden a ir acumulando en el tiempo.

### 2.3.2. *Frame-to-global*

Estas técnicas utilizan el conocimiento del fondo de la escena para ubicar los nuevos *frames* sobre él, y así poder mantener actualizado en tiempo real un fondo completo de la escena, como hacen [35] [36] o [37]. [36] explota el conocimiento del modelo global en conjunto el uso de la propiedad de zoom de las cámaras para obtener información de los parámetros de la cámara y corregir los defectos asociados a las distorsiones originadas por ellos. Obtiene así los parámetros intrínsecos de la cámara, tomando una panorámica al menor nivel de zoom, y posteriormente se determina la distorsión radial incrementando el zoom hasta el máximo posible. Por último, [37] propone crear un PGMM (Panoramic Gaussian Mixture Model) que cubra todo el campo de visión de la cámara, en el que se utiliza un *Gaussian Mixture Model* (aquí previamente mencionado como *MoG*) adaptado a una imagen panorámica previamente creada. Cada nuevo *frame* se inserta en el modelo global y se actualizan los modelos gaussianos.

Para un correcto funcionamiento, este tipo de técnicas necesitan la creación de un fondo inicial sin errores, aunque éstos se den con gran frecuencia.

## Capítulo 3

# Evaluación del sistema base.

El proyecto presentado parte de una motivación relacionada con errores detectados en el sistema de partida o sistema base. Previamente a la propuesta e implementación de soluciones, y con el objetivo de facilitar esta tarea, se va a proceder a confirmar la hipótesis de partida y analizar en detalle las causas y efectos. En esta sección se va a presentar en primer lugar un marco de evaluación sencillo (sección 3.1), con un conjunto de datos completamente orientado a la hipótesis de partida. Posteriormente, y en base a determinadas métricas (definidas en 3.1.2), se va a evaluar el sistema base sobre dicho conjunto de datos (sección 3.2). Como resultado de dicha evaluación se sacarán una serie de conclusiones (sección 3.3) que justificarán las propuestas de solución.

### 3.1. Marco de evaluación.

Los objetivos de la evaluación, y por lo tanto las características que habrá de presentar el conjunto de datos, serán aquellos que se recogieron en la hipótesis de partida.

En dicha hipótesis se defendía que la generación de panorámicas es una tarea compleja en si misma, pero que cuando se realiza sobre entornos controlados se lograba realizar con resultados satisfactorios. También se exponía que cuando esta tarea se llevaba a cabo en entornos no controlados, con cambios de iluminación durante la generación de la panorámica, o con la aparición de elementos de frente ocluyendo el fondo, los resultados obtenidos eran poco satisfactorios.

Con el fin de corroborar la hipótesis, el conjunto de datos habrá de presentar los siguientes objetivos de evaluación: situaciones de frente en movimiento y situaciones de frente estático.

#### 3.1.1. Dataset.

De acuerdo con los objetivos presentados en la subsección anterior, el conjunto de datos o *dataset* ha sido planteado para evaluar el sistema sobre dichos retos. Las secuencias se



Figura 3.1: Situaciones complicadas.

han tomado desde tres cámaras diferentes situadas en el hall de la escuela. Además se ha buscado reflejar tres situaciones: personas en movimiento en una zona cercana a la cámara, en movimiento en una zona alejada de la cámara y paradas durante unos instantes. El objetivo de reflejar dichas situaciones es el siguiente:

- En el caso de personas en movimiento en zonas alejadas de la cámara, se busca evaluar la robustez del sistema ante el movimiento pero con un frente que no ocupa gran cantidad del escenario que está siendo capturado.
- En el mismo caso en zonas más cercanas a la cámara el reto es mayor ya que se añade al movimiento mayor cantidad de frente en cada *frame*.

Adicionalmente, se evalúa el funcionamiento del sistema en secuencias sin frente, esperando un resultado favorable.

La grabación de las secuencias se llevó a cabo con dos personas, aunque se eligió un momento de alta afluencia en la zona, por lo que se aprecian otras personas en movimiento en otras zonas y en diferentes direcciones. Esta situación aporta riqueza de cara a la comprobación del funcionamiento del sistema base. En la figura 3.1 se pueden observar dos de estas situaciones.

El *dataset* consta de un total de 4070 *frames*, con nueve secuencias de una media de 400 *frames* cada una, una resolución de 640 x 480 píxeles y tres situaciones diferentes para cada una de las tres cámaras utilizadas.

#### 3.1.1.1. Nomenclatura

Para identificar cada una de las secuencias se le asigna un nombre de la siguiente forma  $\text{Nombre}_{\text{secuencia}} = \text{Nombre}_{\text{cámara}}\_Situacion$ .

- Nombre de la cámara: CTR. Cámara central; LT. Cámara lateral izquierda; RT. Cámara lateral derecha.



- Situaciones posibles: FAR. Frente en movimiento alejado; CL. Frente en movimiento cercano; ST. Frente estático. FR. Sin frente;

La secuencia de la cámara central con hay frente estático se denomina  $CTR\_ST$ .

### 3.1.1.2. Muestras de las secuencias

A continuación se presentan algunas muestras del conjunto de datos, en la Figura 3.2 , reflejando algunas de las situaciones para las tres cámaras. Para el conjunto completo ver Anexo A.

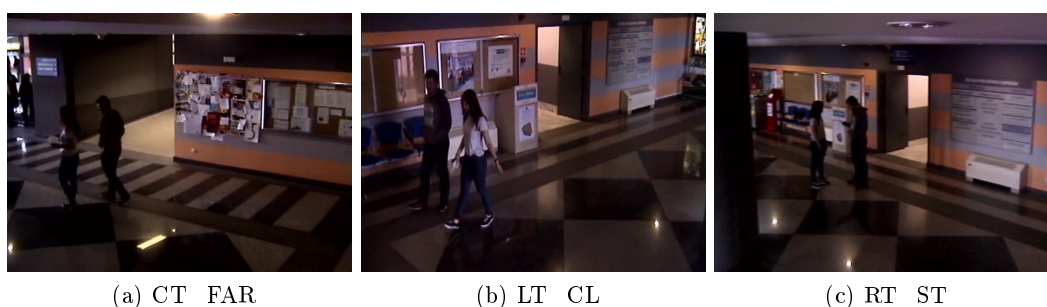


Figura 3.2: Miniaturas del conjunto de secuencias total. De izquierda a derecha, situaciones: frente en movimiento alejado cámara central, frente en movimiento cercano cámara izquierda, frente estático cámara derecha.

Las imágenes de referencia se han generado utilizando información recibida de la propia cámara. En el sistema actual es posible conocer la posición de la cámara en el instante de captura de cada cuadro. Con dicha información, y el conocimiento de la geometría y dimensiones del entorno se han generado imágenes panorámicas perfectas que servirán como base de evaluación (*Ground Truth*). Se muestra un ejemplo en la Figura 3.3.



Figura 3.3: *Ground Truth* para la cámara lateral izquierda

### 3.1.2. Métricas.

Existen numerosas propuestas para evaluar la generación de panorámicas. Sin embargo, el sistema base propuesto se enmarca como primera etapa de otras aplicaciones, en las que

se encarga de la generación de un fondo de escena. Por lo tanto, las métricas escogidas son métricas asociadas a la evaluación de tareas de inicialización de fondo.

Para la selección de métricas se ha recurrido a [38]. En este caso se utilizarán dos: *Average Gray-level Error* (AGE) y *Multi-Scale Structural Similarity Index* (MS-SSIM). A continuación se presenta y comenta brevemente el funcionamiento de cada una de ellas.

### 3.1.2.1. *Average Gray-level Error* (AGE)

Tal como indica el nombre de esta métrica, se calcula el valor absoluto de la diferencia de las dos imágenes, previamente convertidas a escala de grises. La fórmula 3.1 [39] describe su cálculo.

$$AGE = \frac{1}{N} \cdot \sum_{i=1}^N |c(i)_{CB} - c(i)_{GT}| \quad (3.1)$$

La ventaja principal de esta métrica es su extrema sencillez, sin embargo no refleja la diferencia si los objetos de frente son pequeños respecto al tamaño total de la imagen [38]. También tiene problemas en reflejar la igualdad de fondos capturados con distinta iluminación.

### 3.1.2.2. *Multi-Scale Structural Similarity Index* (MS-SSIM)

La métrica SSIM toma como premisa que el sistema visual humano aprecia las formas en una imagen, por lo que se calcula la similitud estructural. Tomando media, varianza y posición de cada píxel se realizan medidas de comparación de luminancia ( $l(x, y)$ ), contraste ( $c(x, y)$ ) y estructural ( $s(x, y)$ ). El índice SSIM se calcula posteriormente como una combinación de estas tres medidas, como se puede ver en 3.2, ponderadas según su relevancia mediante los parámetros  $\alpha$ ,  $\beta$  y  $\gamma$ .

$$SSIM(x, y) = |l(x, y)|^\alpha \cdot |c(x, y)|^\beta \cdot |s(x, y)|^\gamma \quad (3.2)$$

El índice MS-SSIM lleva a cabo esta comparación a diferentes escalas [40], diezmando en un factor 2 cada vez y calculando para cada una de las escalas la comparación estructural y de contraste, aunque no la relativa a la luminancia, para la cual se utiliza solamente la comparación realizada a escala original.

$$MS-SSIM = |l_M(x, y)|^{\alpha_M} \cdot \prod_{j=1}^M |c_j(x, y)|^{\beta_j} \cdot |s_j(x, y)|^{\gamma_j} \quad (3.3)$$

Esta métrica es más fiable para estimar la corrección de los fondos extraídos con diferente iluminación.

## 3.2. Evaluación y análisis

### 3.2.1. Objetivos de la evaluación

Como se ha comentando a lo largo de este capítulo, esta evaluación pretende demostrar el funcionamiento del sistema base, correcto sobre secuencias en las que no hay presencia de frente, pero problemático cuando sí la hay. Por tanto, como se expone en la descripción del *data set* se evalúan cuatro situaciones diferentes: (i) Sin frente; (ii) Con frente en movimiento, alejado de la cámara; (iii) Con frente en movimiento cerca de la cámara; (iv) Con frente estático.

### 3.2.2. Configuración paramétrica

El sistema bajo estudio consta de diferentes etapas, por lo que lógicamente, depende de varios parámetros. No obstante, tan sólo se ha trabajado ajustando un parámetro, denominado *skipped frames*. Este parámetro determina cada cuántos *frames* de la secuencia se procesa uno. Cuando se llevó a cabo el desarrollo del sistema base, [1], se fijó el valor de este parámetro a 20. En consecuencia, para esta evaluación se fija a este valor.

### 3.2.3. Resultados

A continuación se exponen, para cada cámara y según la nomenclatura asignada en la sección 3.1.1.1, los resultados de la evaluación de cada secuencia con las métricas mencionadas en 3.1.2. Se presentan en tres tablas, cada una con las secuencias grabadas para cada cámara.

#### ■ Cámara central

Como se aprecia en los resultados expuestos en 3.1, la puntuación obtenida para esta primera cámara de una panorámica creada sin frente es relativamente mala aunque su apariencia sea muy similar (*Ground Truth* en la figura 3.4a y panorámica generada sin frente en la Figura 3.4b), más si se compara con los resultados obtenidos para las otras dos cámaras. Aún así, los resultados para secuencias en las que se crean situaciones de frente complicadas se encuentran dentro de los márgenes esperados. En este caso la peor puntuación se obtiene para las imágenes en las que hay figuras estáticas durante parte de la secuencia.

Secuencia	AGE	MS-SSIM
CTR_FR	11,5755	0,7206
CTR_FAR	20,8713	0,412
CTR_CL	22,0457	0,4212
CTR_ST	27,2409	0,3665

Tabla 3.1: Tabla resultados cámara central

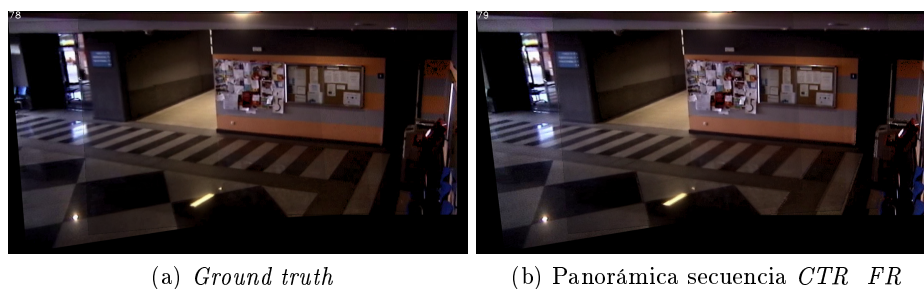


Figura 3.4: Comparativa entre la panorámica tomada como *ground truth* y la panorámica generada a partir de la secuencia *CT\_FR*.

#### ■ Cámara lateral izquierda

En este caso, la similitud entre una panorámica creada a partir de una secuencia sin frente y el *ground truth* es mucho mayor, como se aprecia en 3.2. De nuevo, los peores resultados se obtienen en la situación de frente estático. En la figura 3.5 se muestra la presencia de ruido apreciable para la secuencia *LT\_ST*.

Secuencia	AGE	MS-SSIM
LT_FR	4,8023	0,9216
LT_FAR	16,7287	0,5355
LT_CL	26,9323	0,3662
LT_ST	30,4726	0,3504

Tabla 3.2: Tabla resultados cámara lateral izquierda



Figura 3.5: Panorámica generada para la secuencia *LT\_ST*

#### ■ Cámara lateral derecha

Las secuencias grabadas con ésta cámara son las de mayor complejidad. Este caso pone de manifiesto la poca fiabilidad de la métrica AGE, ya que los resultados de la Tabla 3.3 no se corresponden cualitativamente con las imágenes obtenidas. Como ejemplo, las imágenes de las figuras 3.6a (cuya apariencia se debe al intento de inserción a través de una homografía errónea) y 3.6b.

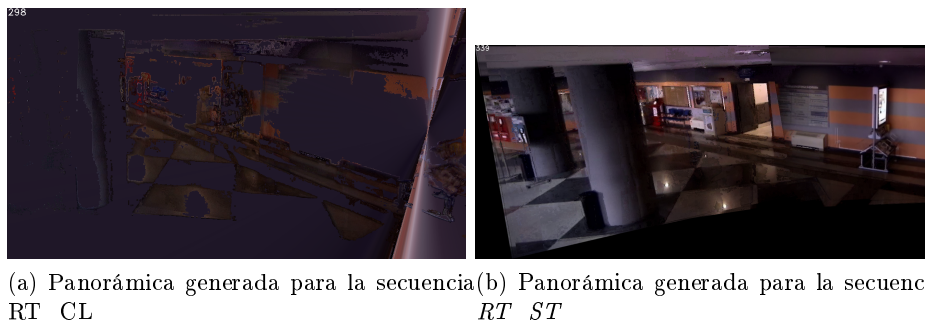


Figura 3.6: Muestra de panorámicas generadas para la cámara lateral derecha

Secuencia	AGE	MS-SSIM
RT_FR	4,1686	0,9375
RT_FAR	10,9208	0,6928
RT_CL	26,5024	0,5219
RT_ST	14,7414	0,5929

Tabla 3.3: Tabla resultados cámara lateral derecha

### 3.2.4. Comparativas

En la figura 3.7 se exponen los resultados para ambas métricas. Si bien no se sigue un patrón muy claro, las secuencias en las que hay frente estático son las que resultan más difíciles de procesar para el sistema base.

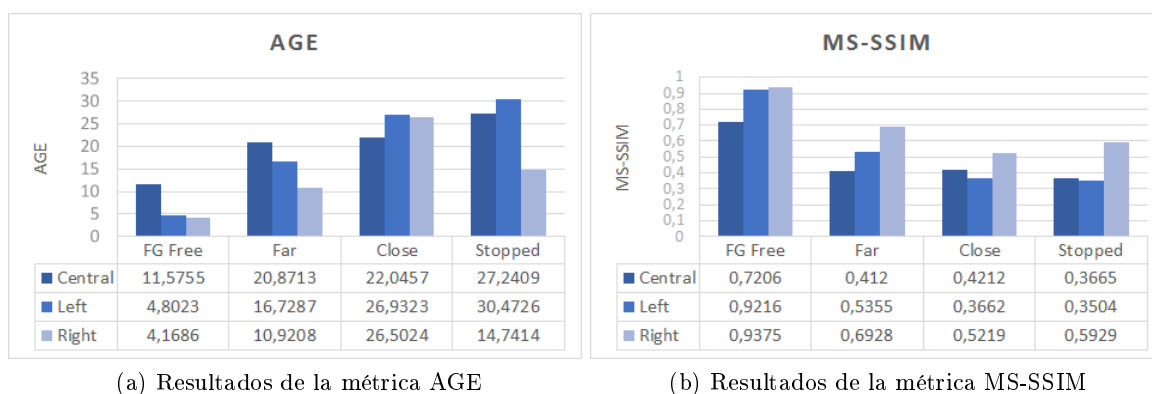


Figura 3.7: Comparativa de los resultados de las métricas según la situación reflejada en cada secuencia

### 3.3. Conclusiones

Tras el análisis de los resultados en 3.2, queda corroborada la hipótesis de la que se partía y se cumplen los objetivos de la evaluación (sección 3.2.1), esto es, que, si bien la creación de panorámicas mediante este sistema se consigue correctamente sin presencia de fondo, los resultados que se obtienen en situaciones más complicadas no son aceptables. Por tanto, en el capítulo siguiente se procede a la descripción de las modificaciones que se proponen para obtener mejores resultados.

## Capítulo 4

# Sistema propuesto.

### 4.1. Sistema base.

El sistema del que se parte tiene como objetivo principal la segmentación frente-fondo. Para ello, en primer lugar se crea una imagen de fondo sobre la que se van comparando las imágenes de entrada. Dadas las características del sistema de captura, una cámara PTZ en movimiento, la imagen de fondo será una panorámica creada al inicio de la grabación.

Para la creación de dicha panorámica se diseñó e implementó un sistema que será la base de la que se parte en este proyecto. Con el objetivo de aclarar la funcionalidad heredada y las modificaciones introducidas, en esta subsección se va a llevar a cabo un breve exposición del sistema de partida.

Para poder describirlo en detalle, se va a introducir una breve nomenclatura inicial.

- Una imagen de entrada recibida de la cámara en un instante temporal  $t$  se define como  $\Psi_t$ .
- La imagen panorámica generada hasta ese instante  $t$  se denominará  $Panorámica_{t-1}$ , y la generada al final de ese instante temporal  $Panorámica_t$ .
- La métrica de asociación de descriptores definida en [28] se nombra como  $\mathcal{M}$ .
- La operación de transformación de una imagen  $\Psi_t$  según una homografía dada  $H_t$ , se representa como  $\mathcal{T}$ .

Una vez expuesta la nomenclatura, el flujo seguido para la generación de la panorámica de fondo es el mostrado en la Figura 4.1. En el Anexo B se van a desarrollar brevemente los módulos principales: puntos SURF y homografía. El módulo de inserción, necesario para desarrollar las modificaciones se explica a continuación.

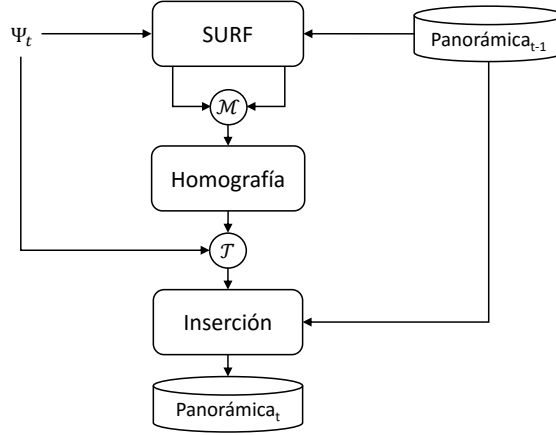


Figura 4.1: Esquema del sistema base para generación de panorámicas.

#### 4.1.1. Inserción

Para poder insertar una nueva imagen,  $\Psi_t$ , sobre una panorámica previa,  $Panorámica_{t-1}$ , es necesario transformar al menos una de las dos a un plano o punto de vista de referencia y común. Con este objetivo se estimó la homografía,  $H_t$ , existente entre  $\Psi_t$  y  $Panorámica_{t-1}$ . Con la información de la homografía disponible el primer paso de esta etapa es la transformación de  $\Psi_t$ . Cabe destacar que se tomó la decisión de realizar un *zero-padding* sobre las zonas que no contienen información tras realizar la transformación.

Una vez se ha generado la imagen transformada,  $\Psi'_t$ , llega el momento de insertarla en  $Panorámica_{t-1}$ . Como criterios de inserción se siguieron dos comunes a las técnicas de generación de fondo de escena: el fondo suele ser la información visual dominante en todas las posiciones de la escena, y el modelo de fondo debe ser lo más ajustado a la realidad posible. Bajo estos criterios se preseleccionaron los métodos de fusión de la media y la mediana. El primero fue descartado dado que generaba un efecto de suavizado en la imagen que era poco deseable según el segundo criterio, por lo que se acabó adoptando la técnica de mediana,  $Me$ , a nivel de píxel sobre las zonas de solape entre la imagen transformada  $\Psi'_t$  y la panorámica hasta ese instante  $Panorámica_{t-1}$ . Como resultado, se obtiene la panorámica en el instante actual  $Panorámica_t = Me(\Psi'_t, Panorámica_{t-1})$ .

## 4.2. Modificaciones introducidas.

Una vez presentado el sistema base, se van a pasar a presentar las modificaciones introducidas. Para ello, ésta subsección va a seguir el siguiente esquema. En primer lugar se presentará el sistema base como un sistema de cajas negras, lo cual simplificará la visualización de las zonas donde se ha trabajado. Este esquema dará pie a una breve exposición de las motivaciones de las modificaciones fundamentadas en el capítulo 3. Por último, se pasará a desglosar en detalle el trabajo realizado en cada una de ellas.



En la Figura 4.2 se muestra en forma de diagrama de flujo de cajas negras el sistema base.



Figura 4.2: Diagrama de flujo de sistema base con cajas negras

El sistema al completo presenta un punto de funcionamiento crítico. Las asociaciones realizadas entre los puntos SURF son la base de todo el proceso, y de no realizarse correctamente el efecto sobre la imagen de fondo generada será muy negativo. Como se demostró en el Capítulo 3, en situaciones donde el fondo se inicializa con elementos de frente ocluyéndolo los resultados del sistema base son muy inferiores a los deseados. Será éste por lo tanto el aspecto que se trate de corregir con las modificaciones introducidas.

La estrategia se dividirá en dos pasos. En primer lugar se implementará un módulo de detección de situaciones conflictivas. Este módulo, compuesto de dos niveles, se denominará módulo de detección de errores. En segundo lugar, el módulo denominado módulo de corrección de errores se encargará de gestionar los errores detectados y tomar decisiones para generar la panorámica final correctamente. En la Figura 4.3 se pueden ver los módulos de modificación introducidos solapados con las distintas etapas del sistema base a las que afectan.

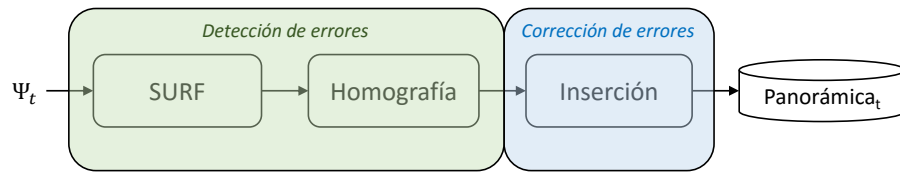


Figura 4.3: Diagrama de flujo de sistema base con bloques de modificación y sus etapas de influencia

La filosofía de las modificaciones propuestas consiste en tratar de aproximar las estrategias multicamino presentadas en el estado del arte, a cámaras móviles. Para ello se habrá de detectar y tratar las zonas, en este caso imágenes de entrada, susceptibles de ser erróneas, y posteriormente se habrá de gestionar la panorámica completa y el recorrido de la cámara de tal manera que se logre asegurar la generación del fondo completo.

### 4.2.1. Módulo de detección de errores

Tal y como se ha expuesto con anterioridad, si se desea evitar la creación de una imagen panorámica errónea, es fundamental que la detección de imágenes que pueden resultar problemáticas se lleve a cabo previamente a su inserción en la panorámica. La detección de imágenes susceptibles de generar errores se puede realizar de diversas formas.

En el estado del arte de la generación de modelos de fondo, y más concretamente en la generación por bloques (como los métodos abordados en las secciones 2.1.3 y 2.1.4 ), que sirve de inspiración a las modificaciones introducidas, esta tarea se enfoca desde el punto de vista de cámaras fijas. En ellas, el nuevo bloque a insertar se compara con los anteriormente insertados, tanto en la misma posición como en las colindantes, obteniendo así una medida de fiabilidad de ser fondo para el bloque de entrada. En estos casos de cámaras estáticas no es necesario realizar ninguna transformación del bloque a insertar antes de realizar la medida. Sin embargo, en la tarea que se aborda en este proyecto si que lo es. Por ello, la propuesta implementará medidas a dos niveles: a nivel previo a la transformación, y a nivel previo a la inserción. Un esquema de las modificaciones se puede observar en la Figura 4.4.

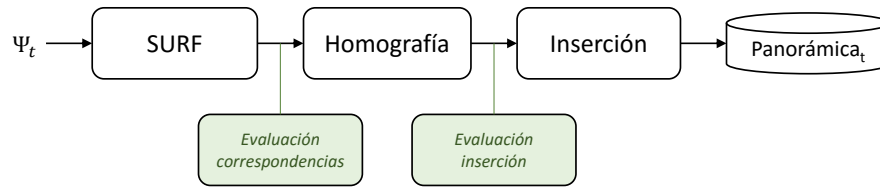


Figura 4.4: Puntos de aplicación del módulo de detección de errores

Con el primer proceso se tratarán de evitar las situaciones en las que gran parte del fondo de la escena quede ocluido por el frente, y por lo tanto las homografías estimadas, y por ende las panorámicas creadas a partir de ahí, sean erróneas. Con el segundo proceso se trata de evitar que elementos de frente sean capturados como elementos de fondo.

#### 4.2.1.1. Evaluación de correspondencias

Tal y como se ha expuesto previamente, los puntos de interés, y más concretamente su correspondencia entre dos imágenes permiten definir objetos o áreas comunes en ambas. Si las correspondencias son bajas o poco fiables, los procesos desarrollados a partir de ellas serán igualmente poco fiables. Para desarrollar el sub-módulo de evaluación de correspondencias fue necesario tener en cuenta dos aspectos.

En primer lugar, determinar cuáles son las situaciones conflictivas o susceptibles de serlo. Para ello, se llevó a cabo una evaluación del sistema base, Capítulo 3, que permitió

determinar que aquellas situaciones en las que el frente en movimiento ocluía grandes áreas del fondo eran las más problemáticas y las que peores resultados generaban.

En segundo lugar, conocer cuáles eran los efectos de estas situaciones sobre las correspondencias extraídas. En este punto es relevante destacar que se podría haber introducido alguna otra métrica de evaluación que no se basase en las correspondencias. Sin embargo no se llevaron a cabo por motivos de buen funcionamiento de las correspondencias como métrica y de no incrementar el coste computacional, aprovechando tareas que ya se realizaban en el sistema base. El efecto observado sobre las correspondencias es el siguiente:

- Detecciones: las técnicas de detección de puntos de interés fijan umbrales relativos a la imagen completa. Al aparecer grandes elementos de frente ocluyendo el fondo, éstos generan gran cantidad de bordes de gran intensidad, por lo que la mayoría de detecciones se centran entorno al elemento de frente.
- Número de asociaciones: los puntos de interés buscan asociar elementos iguales en dos imágenes. Estas técnicas presentan robustez ante pequeñas deformaciones como cambios de iluminación o de punto de vista. Sin embargo, puntos detectados sobre un fondo deformable en movimiento como puede ser una persona, son difícilmente asociados, reduciéndose el número de asociaciones.
- Fiabilidad de las asociaciones: la idea de usar puntos de interés para generar panorámicas se basa en la hipótesis de que éstos son capaces de reflejar el movimiento de la cámara mediante las asociaciones de éstos entre *frames*. Si los puntos son detectados y asociados en su mayoría entre objetos de frente, cuyo movimiento no tiene porque ser coherente con el de la cámara, el movimiento que reflejen será erróneo y por lo tanto la panorámica creada también.

Con estos aspectos controlados, se ha escogido como indicador el número de asociaciones de puntos de interés entre *frames* consecutivos  $t_i$  y  $t_{i+1}$ ,  $\#SURF_{t_i, t_{i+1}}$ , resultantes de la asociación de puntos de interés entre la panorámica  $Panorámica_{t-1}$ , y la imagen de entrada  $\Psi_t$ .

Para llevar a cabo la detección de errores se ha fijado un umbral adaptativo con actualización selectiva,  $\phi$ , que determinará si un nuevo frame es susceptible de provocar errores,  $e$ , o no según 4.1.

$$e = \begin{cases} 1 & \#SURF > \phi \\ 0 & \#SURF < \phi \end{cases} \quad (4.1)$$

El umbral se ha modelado como sigue.

- Inicialización: el valor inicial de umbral,  $\phi_0$ , definido en 4.2, se ha fijado a partir del número de correspondencias obtenidas entre *frames* consecutivos en los instantes

temporales  $\#SURF_{t_1,t_2}$  y  $\#SURF_{t_2,t_3}$ , ponderados por un factor  $\sigma$  definido en según 4.3

$$\phi_0 = \sigma x \frac{\#SURF_{t_1,t_2} + \#SURF_{t_2,t_3}}{2} \quad (4.2)$$

$$\sigma = 1 - \frac{\|\#SURF_{t_1,t_2} - \#SURF_{t_2,t_3}\|}{\#SURF_{t_1,t_2} + \#SURF_{t_2,t_3}} \quad (4.3)$$

- Actualización: el valor del umbral se irá actualizando, añadiendo al cálculo las nuevas cantidades de correspondencias. El valor de  $\sigma$  se mantendrá estable. El valor del umbral de modo genérico para un tiempo  $t'$  se fijaría como se indica en 4.4.

$$\phi = \sigma x \frac{\sum_{i=1}^{t'} \#SURF_{t_i,t_{i+1}}}{t'} \quad (4.4)$$

- Selección: se pretende que el umbral se ajuste a los cambios progresivos, mientras que si sucede una fuerte variación, esos valores no deben usarse para la actualización. Así, los valores que estén por debajo del umbral,  $\phi$ , no se utilizarán en la actualización, y para el resto se utilizará la fórmula genérica 4.4.

En caso de que el indicador evaluado este por debajo del umbral fijado, se activará una señal de error en la evaluación de correspondencias para que el módulo de corrección de errores gestione la alerta.

#### 4.2.1.2. Evaluación de inserción

Tal y como se presentó en la introducción de estas subsecciones, este segundo módulo trata de evitar la inserción de elementos de frente como elementos de fondo. En estas situaciones, muy numerosas, los elementos de frente no suponen un problema para la generación de la panorámica por alguno de los siguientes motivos:

- Fondos lo suficientemente texturados que permiten estimar homografías correctas pese a los elementos de frente.
- Frente estático que permite estimar correctamente las homografías.
- Frente de tamaño reducido, cuyas detecciones pueden tratarse como *outliers*.

En todos estos casos, las imágenes capturadas pueden pasar los filtros de la etapa de evaluación de correspondencias, e incluir información de frente en el fondo panorámico generado. Para tratar de paliar estos efectos, se incluye esta etapa de evaluación de la inserción.

La técnica desarrollada es similar a las propuestas por el estado del arte 2.2.1. La técnica consiste en una métrica que se aplica entre la *Panorámica* <sub>$t-1$</sub>  y la imagen a insertar transformada según la homografía recién estimada,  $\Psi'_t$ . La métrica utilizada, *AGE*, es la misma utilizada para realizar las evaluaciones, que tal y como se definió con anterioridad,

calcula el valor absoluto de la diferencia de las dos imágenes, previamente convertidas a escala de grises.

Se definen dos *sets* de píxeles  $\{c_{\Psi'_t}\}$  y  $\{c_{Pan}\}$ , compuestos  $N_s$  píxeles de las zonas de solape entre la imagen transformada,  $\Psi'_t$ , y la panorámica hasta ese instante,  $Panorámica_{t-1}$ . Sobre ellos se aplica la métrica 3.1 que arroja un valor  $AGE_{in}$ .

Ese valor se umbraliza de nuevo, con respecto a un valor,  $\rho$ , fijado en función del ruido de captura. Para fijar el valor, se aplica 3.1 sobre dos imágenes  $\Psi_t$  y  $\Psi_{t-1}$  capturadas sin movimiento de la cámara en instantes previos a  $t = 1$ . Así se obtiene un valor  $AGE_{th}$ . En función de dicho valor, y según 4.5, se define  $\rho$ .  $\lambda$ , es una ponderación que permite que aquellas muestras que se encuentran dentro de tres veces la desviación típica (99,9 % de las muestras si fuera distribución gaussiana como suele serlo el ruido de la cámara) se consideren como frente, es calculada a partir de 4.6.

$$\rho = AGE_{th} + \lambda \quad (4.5)$$

$$\lambda = 3 \sqrt{\frac{1}{N_s} \cdot \sum_{i=1}^{N_s} (c(i)_{\Psi} - AGE_{th})} \quad (4.6)$$

Al igual que en la subsección anterior, en caso de que durante alguna de las evaluaciones de inserción, no se cumpla el umbral establecido, se activará una señal de error en la evaluación de la inserción para que de nuevo el módulo de corrección de errores actúe en consecuencia.

### 4.3. Módulo de corrección de errores

Tal y como se explica en la sección de evaluación de errores, si se detecta un error se interrumpe el procedimiento normal y se procede a la corrección de errores. Cabe recordar que esta fase solamente tiene lugar si la detección de errores del bloque anterior resulta positiva. La función de este módulo se enfoca sobre la etapa de inserción, ver Figura 4.5.

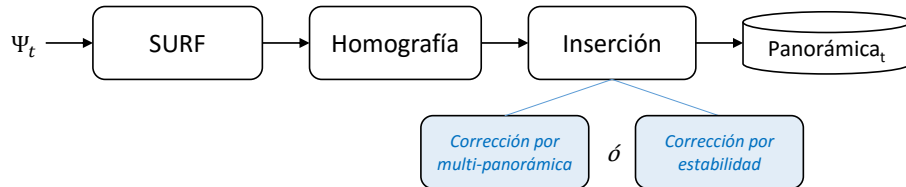


Figura 4.5: Puntos de aplicación del módulo de corrección de errores.

Se van a proponer dos estrategias para proceder con la corrección de errores. La primera de ellas denominada corrección por hipótesis de estabilidad, más conservadora, propone una estrategia basada en la hipótesis de dominancia, esto es, en la hipótesis de que el fondo es la visualización más dominante durante las secuencias de vídeo. La segunda, denominada corrección por multi-panorámica, propone una estrategia más agresiva y compleja, centrándose en un esquema multicamino con rechazo que se detallará a continuación.

#### 4.3.1. Corrección por hipótesis de estabilidad

El esquema de corrección por hipótesis de estabilidad se basa, como se ha comentado anteriormente, en la asunción que llevan a cabo numerosas aproximaciones del estado del arte de inicialización de fondo, de que el fondo es siempre la visualización más repetida.

En base a este criterio, en conjunto con las asunciones hechas al respecto de cuando se producen las detecciones de error, se enuncia la hipótesis sobre la que se sustenta la primera aproximación llevada a cabo como módulo de corrección de errores.

Dicha hipótesis se resume en dos supuestos. En primer lugar, habida cuenta de que las detecciones de situaciones conflictivas se consideran relacionadas directamente con la presencia de elementos de frente en el momento de captura, se considera este supuesto como afirmativo en el momento de llevar a cabo la corrección. En segundo lugar, si el fondo es considerado lo más estable, es razonable considerar que tras una ventana temporal variable, la siguiente visualización de la escena corresponda al fondo y no al frente.

Para desarrollar un módulo que dé soporte a esta hipótesis ha sido necesario fijar un criterio para determinar el tiempo de espera de manera *online*. Dicho tiempo de espera,  $tw$ , se controla mediante una medida,  $t$ , definida en 4.7

$$t = \frac{AGE_{in} - \frac{1}{N_s} \cdot \sum_{i=1}^{N_s} |c(i)_{\Psi'_t} - c(i)_{Pan}|}{AGE_{in}} \quad (4.7)$$

Dicha medida recoge en primer lugar el  $AGE_{in}$  (3.1) del momento de la detección de error y la variación del error en los siguientes cuadros. Cuanto  $t$  cae por debajo del 0,5, esto es, se reduce al menos a la mitad el error en la zona de inserción, se considera que ya se puede reanudar la inserción, y se da por finalizado el tiempo de espera, reseteando su valor. El tiempo de espera está acotado en función de una variable,  $tw_{max}$ , definida en 4.8.

$$tw_{max} = \frac{0,5width}{step} - 1 \quad (4.8)$$

siendo  $step$  el desplazamiento espacial en píxeles de la cámara, obtenido como la variación en el eje x definida por la homografía del frame anterior  $H_{t-1}$ , y  $width$  la anchura también en píxeles del área de solape en el instante de la detección del error.

Con ello, el proceso de control de errores mediante esta técnica queda definido con el diagrama de estados de la Figura 4.6. Si el tiempo de espera cumple alguna de las dos

condiciones, la reducción del error o alcanzar el  $tw_{max}$ , se vuelve al estado de procesado normal. Mientras, el sistema se mantiene en un estado de espera donde únicamente se reciben *frames* y se realizan las medidas indicadas.

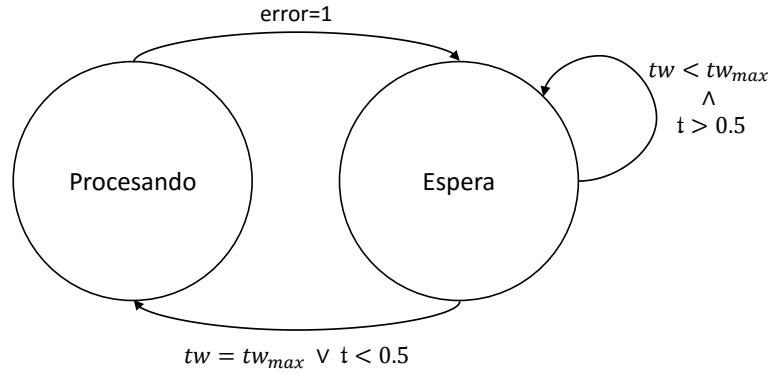


Figura 4.6: Diagrama de estados de la corrección por hipótesis de estabilidad.

#### 4.3.2. Corrección por multi-panorámica

La corrección propuesta en este caso sigue una estrategia similar a las propuestas en el estado del arte en los algoritmos de inicialización de fondo multicamino.

La propuesta pretende capturar en distintas panorámicas el frente y el fondo de la escena, de tal manera que para llevar a cabo la reconstrucción del modelo de fondo solo se utilicen aquellas que contienen la información adecuada, esto es, la información del fondo sin elementos de frente.

Igual que en la aproximación anterior, se parte de dos hipótesis. La primera, igual que en la de estabilidad asume que si se ha detectado una situación de error se debe a la presencia de elementos de frente en la escena. La segunda hipótesis en esta aproximación es que si se genera una panorámica con información de frente, cuando este deje de estar presente se generará una nueva alerta de error, ya que la información entre *frames* consecutivos dejará de ser concordante.

Partiendo de ambas hipótesis se ha diseñado un esquema de generación de nuevas panorámicas con cada señal de error que serán ensambladas, en caso de que sea posible, al final de cada barrido. El funcionamiento se recoge en el esquema mostrado en la Figura 4.7.

Tal y como se ha indicado, en caso de la detección de error se iniciará una nueva panorámica y la anterior se almacenará en un set de panorámicas temporales,  $\{Panorámica_k\}$  para  $k = 1...K$ , siendo  $K$  el número de panorámicas generadas hasta ese momento en el barrido actual. En caso de que se detecte un cambio en la dirección del movimiento de la

cámara, fin de ese barrido, se procederá a ensamblar el set de panorámicas temporales generado en ese barrido. En caso de que ya exista una panorámica general,  $Panorámica_{gral}$ , se tratará de insertar cada elemento de  $\{Panorámica_k\}$  sobre la general como si se tratase de *frames* de entrada individuales. En caso de que no exista panorámica general, se tomará la primera de las panorámicas del set,  $Panorámica_1$ , como la de referencia, y el resto se tratarán de insertar sobre ellas.

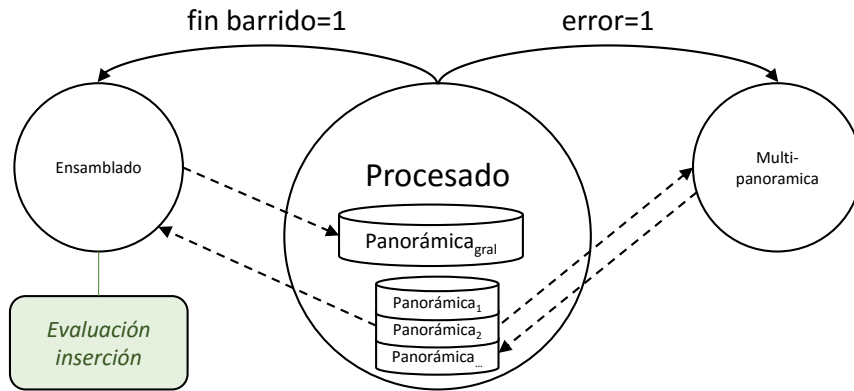


Figura 4.7: Diagrama de estados de la corrección por hipótesis de multi-panorámica.

Por último, la hipótesis multi-panorámica asume que se generarán panorámicas de fondo y de frente. Por ello, se puede observar en la Figura 4.7 que se aplicarán las técnicas de detección de errores, concretamente la de evaluación de la inserción (sección 4.2.1.2). Mediante esta técnica, aquellas panorámicas que den errores de inserción se rechazarán, pues se considerará que serán aquellas que contengan la información de frente. Cabe la posibilidad de que en un solo barrido no se complete la vista completa, sin embargo, se considera que no se debe inicializar el fondo hasta que éste no sea correcto, por lo que será requisito del sistema que todas las partes del fondo hayan sido vistas al descubierto por la cámara al menos en un frame durante toda la secuencia.



## Capítulo 5

# Evaluación del sistema propuesto.

Tras comprobar el mal funcionamiento del sistema base en situaciones exigentes (Capítulo 3) y describir los cambios llevados a cabo para mejorarlo (Capítulo 4), se procede a comprobar si se ha cumplido dicho objetivo. Además, al haberse propuesto dos mejoras diferentes, existe un interés añadido por descubrir cuál de ellas es mejor y en qué casos. Con ese objetivo, el capítulo se va a estructurar como sigue. En primer lugar, se va a fijar el marco de evaluación (sección 5.1) donde se expondrán los objetivos concretos de la misma, el *dataset* empleado y las métricas fijadas. Seguidamente se desarrollarán la evaluación y su correspondiente análisis (sección 5.2), con la explicación del ajuste paramétrico, los resultados desglosados y la comparativa. Se cerrará el capítulo con una serie de conclusiones (sección 3.3) particulares sobre la evaluación.

Por simplicidad, en este capítulo se hace referencia a las dos propuestas mencionadas en el capítulo anterior de la siguiente forma:

- **HE** hace referencia a la implementación del módulo de corrección de errores mediante **hipótesis de estabilidad** (sección 4.6).
- **MP** hace referencia a la implementación del módulo de corrección de errores mediante **multi-panorámica** (sección 4.3.2).

### 5.1. Marco de evaluación.

Esta evaluación viene a corroborar que las mejoras introducidas en el sistema presentan el funcionamiento que cabía esperar. Adicionalmente se espera también comprobar el funcionamiento del sistema en situaciones normales, donde los retos evaluados en el Capítulo 3 pueden o no estar presentes. Para ello, se va a presentar un conjunto de datos de evaluación más completo, una métrica que complete el conjunto de las anteriormente utilizadas, y una evaluación que permita sacar las conclusiones buscadas.

### 5.1.1. *Dataset.*

Tal y como se ha mencionado anteriormente, el conjunto de datos utilizado para evaluar la mejora incluye aquellos descritos en el Capítulo 3. De forma adicional, se han añadido tres secuencias, grabadas con una de las cámaras que también se utilizaron para el *dataset* original. En estas secuencias se reflejan situaciones menos exigentes para el sistema de generación de panorámicas. Estas nuevas secuencias tienen un *frame rate* de 20 fps, diferente al de las secuencias del *dataset* anterior (10 fps).

Por tanto, al *dataset* anterior se añaden 1020 *frames*, que se distribuyen en tres nuevas secuencias en diferentes situaciones para una misma cámara, con una resolución de 580 x 300 píxeles. Esto da lugar a un *dataset* con un total de 5090 *frames*.

Las nuevas imágenes siguen una nomenclatura de acuerdo a la descrita en 3.1.1.1.

### 5.1.2. *Métricas.*

Para evaluar cuantitativamente este sistema, además de las métricas usadas en la sección 3.1.2, se utiliza una nueva métrica, *Percentage of Error Pixels* (pEPs). Esta decisión se ha tomado debido a las carencias del método AGE. Un ejemplo muy ilustrativo es el mostrado en la Figura 5.1, en el que ambas imágenes difieren décimas en la puntuación dada por AGE, pese a que visualmente el error de una y otra es claramente diferenciable. Por contra, pEPs, junto a MS-SSIM, refleja el resultado cuantitativo de manera mucho más ajustada a la percepción visual [38].

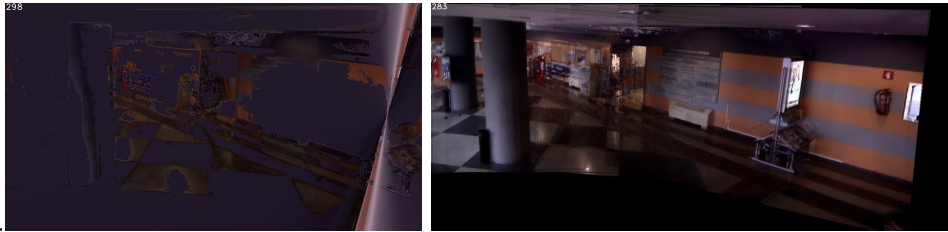


Figura 5.1: Panorámica creada para la misma secuencia con el sistema base (izquierda) y con una de las soluciones propuestas (derecha).

#### 5.1.2.1. *Percentage of Error Pixels (pEPs)*

Se denomina *error pixel* a aquel cuyo valor difiere más de un valor determinado del mismo pixel en la imagen *ground truth*. Esta métrica mide el porcentaje de píxeles totales erróneos mediante el cociente entre estos y la cantidad total de píxeles en la imagen bajo evaluación, como se puede ver en 5.1, siendo *EPs* el número de píxeles erróneos y *N* el número total de píxeles.

$$pEps = \frac{EPs}{N} \quad (5.1)$$

## 5.2. Evaluación y análisis

### 5.2.1. Ajuste de evaluación

Al igual que ocurría en el sistema base, en este sistema existen parámetros cuyo valor inicial puede ser necesario fijar. Sin embargo, tal y como se presentó en las secciones 4.2.1 y 4.3, se ha pretendido que todos los nuevos parámetros sean autoajustables, por lo que para ellos no habrá que fijar dicho valor inicial. Con esto, el único parámetro que el sistema requiere fijar manualmente es el heredado del sistema base, *skipped frames*. Con el objetivo de analizar su influencia en el comportamiento del sistema, se barrerán los valores de 10, 20 y 40, es decir, aquel que se consideró para la implementación inicial del sistema base, la mitad de dicho valor y el doble.

### 5.2.2. Resultados

A continuación se presentan los resultados obtenidos de la evaluación de todas las panorámicas obtenidas con los diferentes sistemas. La estructura que sigue la sección es similar a la que se presenta en la sección 3.2.3. En este caso, se presentan los resultados para cada cámara mediante gráficos, uno para cada una de las métricas mencionadas en 5.1.2 para todas las secuencias del *dataset*. Estos datos se presentan de forma más detallada, mediante tablas, en el anexo C.

Cabe recordar que, mientras que para las métricas AGE (3.1.2.1) y pEPs (5.1.2.1) se consideran mejores resultados bajos valores, la métrica MS-SSIM (3.1.2.2) indica similitud, por lo que cuanto mayor sea su valor mejor es el resultado.

#### 5.2.2.1. Resultados MP

Como bien se describe en la sección 4.3.2, este sistema comienza una panorámica diferente cuando detecta la posibilidad de *frames* conflictivos (indicador de detección de errores). A continuación se exponen los resultados de las métricas sobre las panorámicas generadas por este sistema.

##### ■ Cámara central

Como puede apreciarse en la figura 5.2, los resultados para cada uno de los valores de *skipped frames* apenas difieren, si bien llama la atención la gran diferencia entre los resultados obtenidos para la secuencia CTR\_FAR cuando el valor de este parámetro se fija a 10, 20 y 40. Al igual que en la evaluación del sistema base, la peor puntuación se obtiene en el caso de frente estático.

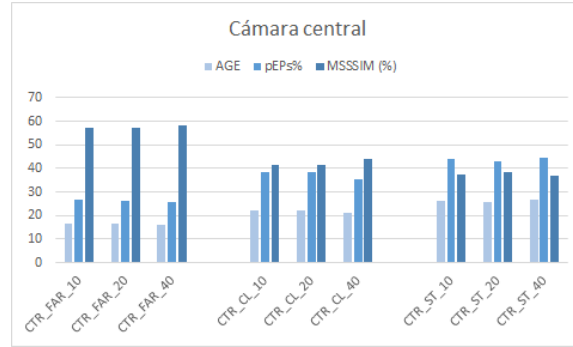


Figura 5.2: Resultado de las métricas para la cámara central

#### ■ Cámara lateral izquierda

En este caso, de nuevo, el peor caso ocurre en el caso de frente estático. Comparando entre valores para skipped frames, según se puede ver en la Figura 5.3, los resultados obtenidos cuando este parámetro es 40 son ligeramente mejores que para 10 y 20.

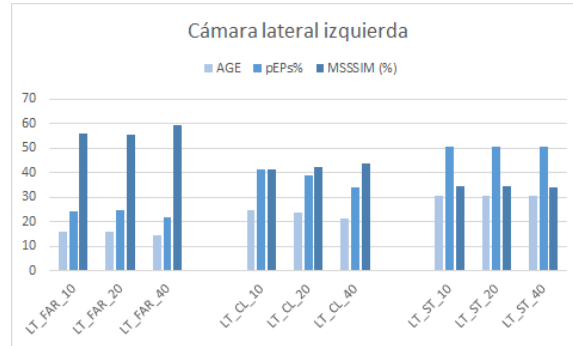


Figura 5.3: Resultado de las métricas para la cámara lateral izquierda

#### ■ Cámara lateral derecha

Para esta cámara los resultados son menos homogéneos, pero en este caso los resultados cuando el número de *skipped frames* es 40 son peores que cuando es 10 o 20. Puede apreciarse en la figura 5.4

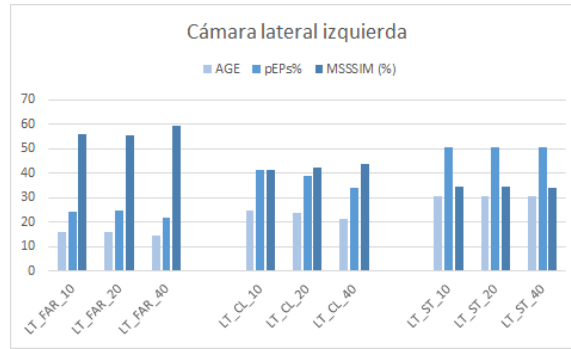


Figura 5.4: Resultado de las métricas para la cámara lateral derecha

#### 5.2.2.2. Resultados HE

Como se describe en 4.3.1, este otro sistema se basa en, al constatar la presencia de frente, esperar un número determinado de *frames* para seguir procesando *frames* e insertándolos en la panorámica final.

##### ■ Cámara central

En este caso se obtienen resultados muy dispares (Figura 5.5 ),de forma que no destaca ninguno de los valores de *skipped frames* sobre los otros.

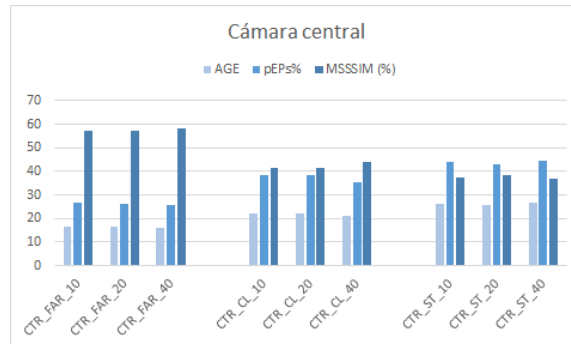


Figura 5.5: Resultado de las métricas para la cámara central

##### ■ Cámara lateral izquierda

De forma global, según se aprecia en la Figura 5.6, en este caso el valor de *skipped frames* que mejor funciona es 20 , aunque para el valor 10 se obtiene mejor resultado para la situación en la que hay frente estático.

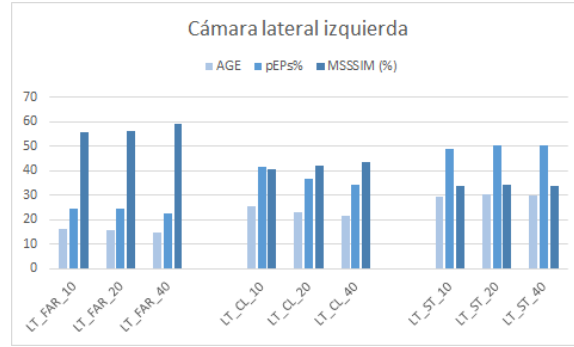


Figura 5.6: Resultado de las métricas para la cámara lateral izquierda

#### ■ Cámara lateral derecha

Con estos resultados se vuelve a constatar la singularidad de los resultados (figura 5.7) obtenidos para las secuencias grabadas en esta cámara. En este caso la imagen que peor puntuación se obtiene es *RT\_CL*, aunque es previsible por lo descrito en la sección 3.1.1.

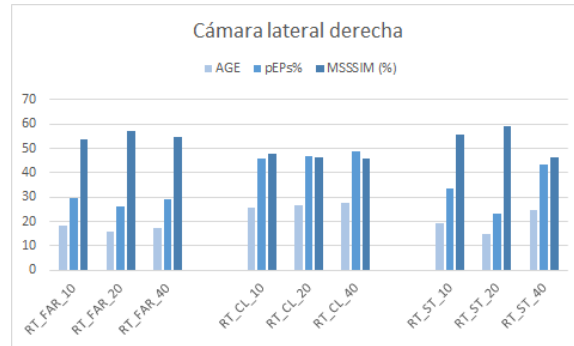


Figura 5.7: Resultado de las métricas para la cámara lateral derecha

#### 5.2.2.3. Sistema base

Puesto que ya se ha analizado en profundidad el comportamiento de este sistema en el Capítulo 3, las gráficas correspondientes a los resultados obtenidos en este caso se encuentran en el Anexo C. En este caso se presentan también los resultados obtenidos para los tres valores de *skipped frames*, para poder hacer una comparación más exhaustiva de este sistema y los propuestos. Sí cabe destacar que este sistema no ha podido generar una panorámica para la secuencia *RT\_CL*. En su lugar se indica un 0 en las métricas.

	AGE	pEPs	MS-SSIM	Total	Total %
Sistema base	8	10	9	27	36
Sistema MP	17	15	16	48	64

(a) Comparativa sistema base vs sistema MP

	AGE	pEPs	MS-SSIM	Total	Total %
Sistema base	8	10	11	29	39
Sistema HE	17	15	14	46	61

(b) Comparativa sistema base vs sistema HE

Tabla 5.2: Comparativa individual de cada sistema propuesto con el sistema base

### 5.2.3. Comparativas

#### 5.2.3.1. Sistema base vs sistemas propuestos

##### Cuantitativas

Debido a la gran cantidad de datos recogidos, no resulta fácil llevar a cabo una comparativa de forma directa. A continuación se muestra, en la Tabla 5.1, el número de secuencias para las cuales un sistema ha obtenido mejor puntuación que los otros dos, y el porcentaje que este representa del total.

	AGE	pEPs	MS-SSIM	Total	Total %
Sistema base	6	8	8	22	23
Sistema MP	7	7	11	25	26
Sistema HE	14	12	8	48	51

Tabla 5.1: Número de veces que cada sistema ha obtenido la mejor puntuación

Estos datos demuestran que, utilizando alguno de los sistemas propuestos, los resultados mejoran en un 77 % de los casos respecto a los del sistema base. Además, se aprecia que, para la mitad de las secuencias analizadas para los tres sistemas, la mejor puntuación es obtenida por la segunda mejora. En una comparativa individual de cada uno de los sistemas propuestos con el sistema base, como se hace en la Tabla 5.2, tienen gran ventaja de nuevo los sistemas propuestos (con unos resultados muy similares).

##### Cualitativas

Puesto que en algunos casos las medidas numéricas no reflejan realmente los resultados obtenidos, a continuación se presentan las panorámicas generadas por los diferentes sistemas

para la misma secuencia (Figura 5.8). La mayor diferencia se encuentra en el ruido que aparece en la zona central de la imagen, en mayor o menor medida según con qué sistema haya sido la panorámica. Para apreciar mejor estas diferencias, las imágenes de la figura se encuentran, en mayor tamaño, en el anexo D.

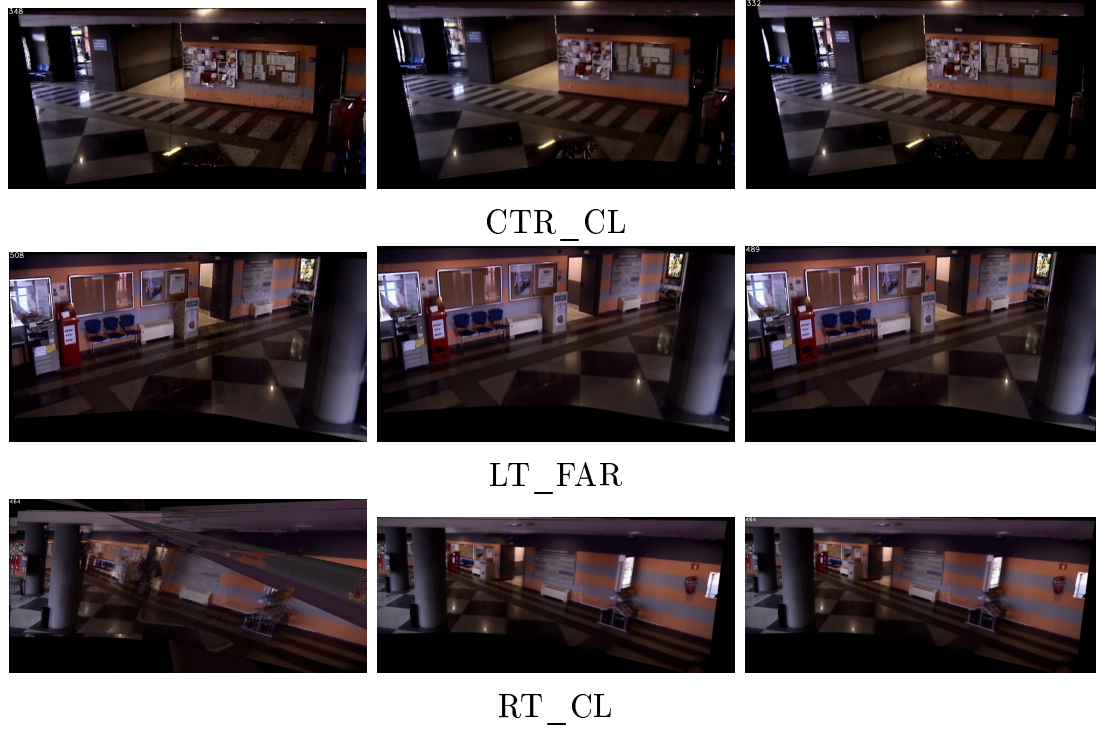


Figura 5.8: Comparativa de panorámicas generadas para la misma secuencia con los diferentes sistemas: de izquierda a derecha, sistema base, sistema MP y sistema HE

#### 5.2.3.2. Sistema MP vs sistema HE

En la sección anterior las evaluaciones son más favorables para el sistema HE. No obstante, en esta sección se expone la comparación de las secuencias evaluadas exclusivamente con estos dos sistemas y se remarcan los puntos fuertes de cada uno de ellos.

**Comparativa cuantitativa** Igual que en la sección 5.2.3, la comparación se hará basándose en cuál de los sistemas ha obtenido mejores resultados en todas las métricas obtenidas del *dataset* completo descrito en 5.1.1. Los resultados se exponen en la tabla. Curiosamente, en este caso, el sistema MP obtiene más veces la mejor puntuación que el sistema HE.



	AGE	pEPs	MS-SSIM	Total	Total (%)
Sistema MP	21	19	18	58	54
Sistema HE	15	17	18	50	46

Tabla 5.3: Número de veces que cada sistema ha obtenido la mejor puntuación

**Comparativa cualitativa** De nuevo, la comparación de resultados resulta más ilustrativa si se muestra visualmente, como en la figura 5.9. El problema que sufre el sistema MP es que, al basarse en la superposición de varias panorámicas, este efecto se aprecia en el resultado final, como se puede ver en la panorámica final para la secuencia SEQ\_FAR. Sin embargo, para panorámicas más complejas, como RT\_CL, se consigue eliminar mayor cantidad de ruido que el sistema HE.



Figura 5.9: Comparativa de panorámicas generadas con los dos sistemas propuestos

### 5.3. Conclusiones

Las evaluaciones llevadas a cabo en este capítulo tenían dos objetivos: analizar el comportamiento del sistema variando el parámetro *skipped frames* (mencionado en la sección 5.2.1) y comparar los dos sistemas propuestos con el sistema base.

Respecto al análisis del valor del parámetro, se obtienen mejores resultados con el mayor valor que se ha dado, 40. Esto se debe a que cuanto mayor es el número de *frames* que transcurre entre dos *frames* procesados, si hay presencia de frente éste será más fácil de detectar, y por tanto el sistema actuará en consecuencia. El problema de aumentar el valor de este parámetro es que hay peligro de no procesar todas las zonas de la panorámica, dependiendo de la longitud de la secuencia que se esté procesando.

A la vista de los resultados obtenidos en las comparativas, secciones 5.2.3.1 y 5.2.3.2, se pueden extraer varias conclusiones:

- En primer lugar, que el sistema base ha sido mejorado. Las puntuaciones en las evaluaciones del sistema HE superan casi en su totalidad las del sistema base, y las del

sistema MP también las mejoran en varios casos. Aún así, es en las situaciones más conflictivas donde se aprecia visualmente la verdadera mejora en las panorámicas generadas por cada sistema (5.8). Cuando la cantidad de frente es considerable, en las panorámicas que han sido generadas con el sistema base aparece mucho más ruido, que en los sistemas propuestos es menor o inexistente.

- En segundo lugar, que el sistema HE arroja mejores resultados, numéricamente hablando, que el sistema MP. Al comparar visualmente las panorámicas que han sido creadas, se aprecia que estas bajas puntuaciones se ven provocadas por la superposición de varias panorámicas de diferentes tamaños. Sin embargo, en las situaciones más exigentes, se consiguen resultados apreciablemente mejores, al menos en la zona central. En los laterales, como ya se ha comentado previamente, surgen problemas al intentar unir varias panorámicas de diferentes tamaños. Esto crea “múltiples bordes”, claramente apreciables en determinadas panorámicas. Sin embargo, en la zona de la imagen en la que ha habido frente, éste se aprecia poco o nada.

En conclusión, el objetivo de mejorar la creación de panorámicas del sistema base se ha cumplido. Respecto a los sistemas desarrollados, se proponen para diferentes situaciones: el sistema HE para situaciones menos exigentes, y el sistema MP para situaciones en las que el frente supone un gran reto (como, por ejemplo, en zonas altamente transitadas).

## Capítulo 6

# Conclusiones y trabajo futuro.

### 6.1. Conclusiones

El objetivo de este trabajo de fin de grado era el de intentar solucionar los problemas del *Background Initialization* dadas las condiciones de cámara móvil e “inicio en caliente”. Para ello, se ha desarrollado un sistema que detecta la presencia de situaciones susceptibles de generar modelos de fondo (panorámicas) erróneos y propone una serie de estrategias para paliar los efectos de estas situaciones.

En primer lugar, se detectaron las situaciones en las que el sistema bajo prueba fallaba. Para ello se evaluó el sistema sobre un conjunto de datos construido en base a una serie de hipótesis de situaciones conflictivas para el sistema base. Fijado el conjunto de datos, se llevó a cabo una evaluación que permitió no solo corroborar las hipótesis de fallo, sino analizar los efectos sobre los distintos elementos del sistema para proponer soluciones en consecuencia.

Posteriormente, y en base al análisis realizado, se propusieron varias mejoras del sistema de partida. Como ya se ha comentado, la estrategia para mejorar el sistema consistió en detectar las situaciones conflictivas, habitualmente relacionadas con la presencia de elementos de frente, e intentar evitar que estas quedasen reflejadas en el resultado final en la medida de lo posible. Para la detección se llevó a cabo la evaluación de correspondencias utilizadas en el sistema base para generar las homografías entre imágenes. Para la corrección de los errores que se detectaban, se optó por desarrollar dos estrategias, de diferente complejidad, que proponían soluciones relacionadas con la estabilidad del fondo y las estrategias multicamino con rechazo.

Por último, para corroborar el funcionamiento de las mejoras y comprobar en qué situaciones funcionaba mejor cada una de las dos, se completó el *dataset* anteriormente propuesto, incorporando algunas secuencias en condiciones de normalidad. A partir de la evaluación realizada se pudo concluir que las propuestas generaban soluciones con mejores resultados tanto cuantitativos como cualitativos.

Como conclusión final del proyecto, se puede decir que se han conseguido los objetivos, mejorando de manera apreciable los resultados del sistema de partida.

Más allá de la consecución de los objetivos, es posible extraer una serie de conclusiones como resultado de los distintos problemas abordados durante la realización del trabajo.

En primer lugar, la idea inicial de mejora, según la cual la panorámica final se consigue mediante *stitching* de panorámicas de menor tamaño, arrojaba resultados nada alentadores en un principio. Esto fue debido al pobre funcionamiento de los puntos de interés a la hora de la estimación de homografías entre imágenes donde los cambios de punto de vista son notables.

Por otra parte, las evaluaciones, y más concretamente las métricas, han llevado en ocasiones a realizar lecturas erróneas de lo que realmente estaba sucediendo. En primer lugar, no todas las métricas reflejan con fidelidad las diferencias apreciables a simple vista entre modelos de fondo generados erróneamente y modelos generados correctamente. En segundo lugar, para realizar unas medidas lo más fiables posibles es necesario alinear las imágenes resultado y las de *ground truth*. Sin embargo, esto no siempre es posible, principalmente si las imágenes se generan deformadas o de distintas dimensiones debido a los propios errores. Por lo tanto surgen ciertas reticencias respecto a la diferencia real entre los resultados del sistema propuesto y el sistema de partida.

## 6.2. Trabajo futuro

El sistema desarrollado ha surgido y se ha implementado exclusivamente en cámaras PTZ. Sin embargo, su diseño hace pensar que sería totalmente compatible con otro tipo de cámaras móviles. En la misma línea, esto permitiría probar dicho sistema en situaciones de mayor libertad, ya que las cámaras PTZ se encuentran ancladas a un punto fijo.

A partir del sistema que se ha desarrollado son posibles varias aplicaciones: por ejemplo, la reconstrucción de escena a partir de una secuencia en la que se graban partes inconexas de la escena o de forma diferente al barrido de lado a lado convencional. También sería posible la detección de objetos abandonados o cambios en una escena mediante la comparación de imágenes generadas en diferentes momentos.





# Bibliografía

- [1] A. Palero, “Detección de intrusión con cámaras móviles en tiempo real. trabajo fin de grado, eps, uam,” Master’s thesis, 2014.
- [2] Á. Bayona, J. C. SanMiguel, and J. M. Martínez, “Stationary foreground detection using background subtraction and temporal difference in video surveillance,” in *2010 IEEE International Conference on Image Processing*, pp. 4657–4660, IEEE, 2010.
- [3] F. N. Fajardo, M. Escudero-Viñolo, and J. B. Cano, “Enhancing region-based object tracking with the sp-sift feature,” in *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 1–4, IEEE, 2014.
- [4] A. Shio and J. Sklansky, “Segmentation of people in motion,” in *Visual Motion, 1991., Proceedings of the IEEE Workshop on*, pp. 325–332, IEEE, 1991.
- [5] H. Eng, K. Toh, A. H. Kam, J. Wang, and W.-Y. Yau, “An automatic drowning detection surveillance system for challenging outdoor pool environments,” in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 532–539, IEEE, 2003.
- [6] B. Gloyer, H. K. Aghajan, K.-Y. Siu, and T. Kailath, “Video-based freeway-monitoring system using recursive vehicle tracking,” in *IS&T/SPIE’s Symposium on Electronic Imaging: Science & Technology*, pp. 173–180, International Society for Optics and Photonics, 1995.
- [7] B. Lo and S. Velastin, “Automatic congestion detection system for underground platforms,” in *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on*, pp. 158–161, IEEE, 2001.
- [8] M. Massey and W. Bender, “Salient stills: Process and practice,” *IBM Systems Journal*, vol. 35, no. 3.4, pp. 557–573, 1996.
- [9] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, IEEE, 1999.
- [10] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” in *Computer vision–ECCV 2000*, pp. 751–767, Springer, 2000.
- [11] L. Maddalena and A. Petrosino, “A self-organizing approach to background subtraction for visual surveillance applications,” *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1168–1177, 2008.
- [12] W. Long and Y.-H. Yang, “Stationary background generation: An alternative to the difference of two images,” *Pattern Recognit.*, vol. 23, no. 12, pp. 1351–1359, 1990.

- [13] D. Gutchess, M. Trajković, E. Cohen-Solal, D. Lyons, and A. K. Jain, "A background model initialization algorithm for video surveillance," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, pp. 733–740, IEEE, 2001.
- [14] A. J. Lipton and N. Haering, "Commode: An algorithm for video background modeling and object segmentation," in *Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference on*, vol. 3, pp. 1603–1608, IEEE, 2002.
- [15] A. Colombari, M. Cristani, V. Murino, and A. Fusiello, "Exemplar-based background model initialization," in *Proceedings of the third ACM international workshop on Video surveillance & sensor networks*, pp. 29–36, ACM, 2005.
- [16] H.-H. Lin, T.-L. Liu, and J.-H. Chuang, "Learning a scene background model via classification," *Signal Processing, IEEE Transactions on*, vol. 57, no. 5, pp. 1641–1654, 2009.
- [17] D. Baltieri, R. Vezzani, and R. Cucchiara, "Fast background initialization with recursive hadamard transform," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pp. 165–171, IEEE, 2010.
- [18] X. Xu and T. S. Huang, "A loopy belief propagation approach for robust background estimation," in *Computer Vision and Pattern Recognition CVPR. IEEE Conference on*, pp. 1–7, IEEE, 2008.
- [19] X. Chen, Y. Shen, and Y. H. Yang, "Background estimation using graph cuts and inpainting," in *Proceedings of Graphics Interface*, pp. 97–103, Canadian Information Processing Society, 2010.
- [20] T. Crivelli, G. Piriou, P. Bouthemy, B. Cernuschi-Frías, and J.-f. Yao, "Simultaneous motion detection and background reconstruction with a mixed-state conditional markov random field," in *Computer Vision—ECCV*, pp. 113–126, Springer, 2008.
- [21] C. Guo, S. Gao, and D. Zhang, "Belief propagation algorithm for background estimation based on local maximum weight matching," in *Image and Signal Processing (CISP), 5th International Congress on*, pp. 82–85, IEEE, 2012.
- [22] D. Park and H. Byun, "A unified approach to background adaptation and initialization in public scenes," *Pattern Recognition*, vol. 46, no. 7, pp. 1985–1997, 2013.
- [23] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," *Exploring artificial intelligence in the new millennium*, vol. 8, pp. 236–239, 2003.
- [24] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.
- [25] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006.
- [26] B. D. Lucas, T. Kanade, *et al.*, "An iterative image registration technique with an application to stereo vision," in *IJCAI*, vol. 81, pp. 674–679, 1981.
- [27] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.



- [28] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision-ECCV 2006*, pp. 404–417, Springer, 2006.
- [29] E. Oyallon and J. Rabin, "An analysis of the surf method," *Image Processing On Line*, vol. 5, pp. 176–218, 2015.
- [30] N. Martinel, C. Micheloni, and G. L. Foresti, "Robust painting recognition and registration for mobile augmented reality," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1022–1025, 2013.
- [31] N. Liu, H. Wu, and L. Lin, "Hierarchical ensemble of background models for ptz-based video surveillance," *Cybernetics, IEEE Transactions on*, vol. 45, no. 1, pp. 89–102, 2015.
- [32] S. Kang, J.-K. Paik, A. Koschan, B. R. Abidi, and M. A. Abidi, "Real-time video tracking using ptz cameras," in *Quality Control by Artificial Vision*, pp. 103–111, International Society for Optics and Photonics, 2003.
- [33] S. Wu, T. Zhao, C. Broaddus, C. Yang, and M. Aggarwal, "Robust pan, tilt and zoom estimation for ptz camera by using meta data and/or frame-to-frame correspondences," in *Control, Automation, Robotics and Vision, 2006. ICARCV'06. 9th International Conference on*, pp. 1–7, IEEE, 2006.
- [34] M. Irani and P. Anandan, "A unified approach to moving object detection in 2d and 3d scenes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 6, pp. 577–589, 1998.
- [35] P. Azzari, L. D. Stefano, and A. Bevilacqua, "An effective real-time mosaicing algorithm apt to detect motion through background subtraction using a ptz camera," in *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*, pp. 511–516, IEEE, 2005.
- [36] S. N. Sinha and M. Pollefeys, "Pan-tilt-zoom camera calibration and high-resolution mosaic generation," *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 170–183, 2006.
- [37] K. Xue, Y. Liu, G. Ogunmakin, J. Chen, and J. Zhang, "Panoramic gaussian mixture model and large-scale range background subtraction method for ptz camera-based surveillance systems," *Machine vision and applications*, vol. 24, no. 3, pp. 477–492, 2013.
- [38] L. Maddalena and A. Petrosino, "Towards benchmarking scene background initialization," in *International Conference on Image Analysis and Processing*, pp. 469–476, Springer, 2015.
- [39] C.-h. Chen, *Computer vision in medical imaging*, vol. 2. World scientific, 2014.
- [40] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, pp. 1398–1402, Ieee, 2003.



## Apéndice A

### Dataset evaluación sistema base

A continuación se muestran las miniaturas correspondientes a las secuencias incluidas en el conjunto de datos utilizado para la evaluación del sistema base.



Figura A.1: Miniaturas del conjunto de secuencias total. Por filas, cámaras: central, izquierda y derecha. Por columnas, situaciones: frente en movimiento alejado, frente en movimiento cercano, frente estático.

## Apéndice B

# Módulos del sistema base

### B.0.1. Puntos SURF

En la sección 2.2.2 se analizaron diversas técnicas para estimar la posición de un cuadro con respecto a otro cuando estos comparten áreas de visión. Entre todas las expuestas, aquella que garantizaba mejor balance entre coste computacional y resultados era la basada en puntos de interés. A esta misma conclusión se llegó cuando se desarrolló el proyecto base, por lo que fue una técnica de puntos la desarrollada para llevar a cabo la generación de la imagen panorámica de fondo.

Tras un breve estudio del estado del arte de las técnicas de puntos de interés realizado en el proyecto original, se llegó a la conclusión de que los puntos SURF eran la mejor opción. La decisión se fundamentó tanto por coste computacional como por resultados, así como por tratarse de una referencia del estado del arte suficientemente contrastada.

El proceso de detección y descripción sobre puntos detectados independientemente sobre la imagen de entrada  $\Psi_t$  y sobre  $Panorámica_{t-1}$ . Una vez se han detectado y descrito los puntos SURF correspondientes a cada imagen se dispone de dos sets de vectores de descripción  $\{\vec{f}_\Psi\}$  y  $\{\vec{f}_{Pan}\}$ . Sobre dichos sets se aplica la métrica de asociación  $\mathcal{M}$  originando un conjunto de asociaciones entre los descriptores de ambos sets, y por ende de sus coordenadas asociadas  $\{(x, y)_\Psi, (x, y)_{Pan}\}$ .

### B.0.2. Homografía

Una homografía describe la transformación que experimenta un plano proyectado cuando la posición del observador (o de la cámara, en este caso) varía. En el caso de las cámaras PTZ, ancladas a un punto fijo, varía es el ángulo desde el cual se toma la imagen. La homografía toma forma de matriz que transforma la imagen original a la perspectiva deseada de la forma descrita en B.1, siendo  $\vec{x} = \begin{bmatrix} x & y & z \end{bmatrix}$  las coordenadas de un píxel de la imagen a transformar y  $\vec{x'} = \begin{bmatrix} x' & y' & z' \end{bmatrix}$  sus coordenadas correspondientes en la imagen transformada.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (\text{B.1})$$

En este caso, la homografía se establece haciendo uso de la asociación de coordenadas obtenida en B.0.1.

## Apéndice C

# Resultados de evaluación

Las tablas que aquí se presentan corresponden a los resultados de aplicar las métricas indicadas en el Capítulo 6.

### C.0.0.1. Resultados Multi-Panorámica

#### ■ Cámara central

Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
CTR_FAR	27,9998	49,2955	0,3309	CTR_FAR	19,5659	32,5208	0,4887
CTR_CL	21,8095	36,3813	0,4477	CTR_CL	21,5465	36,5835	0,4381
CTR_ST	25,87	43,5114	0,3817	CTR_ST	25,9259	43,2143	0,3843

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
CTR_FAR	15,7762	25,518	0,5723
CTR_CL	20,389	33,8414	0,4426
CTR_ST	26,4223	43,9611	0,3697

(c) Skipped frames = 40

Tabla C.1: Resultados para la cámara central

#### ■ Cámara lateral izquierda

Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
LT_FAR	15,8164	24,2764	0,562	LT_FAR	16,0041	24,8947	0,5553
LT_CL	24,939	41,2282	0,4121	LT_CL	23,5821	38,748	0,422
LT_ST	30,7218	50,5445	0,3476	LT_ST	30,4334	50,5952	0,3441

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
LT_FAR	14,6401	22,0177	0,5969
LT_CL	21,1727	34,117	0,4398
LT_ST	30,4045	50,6718	0,3393

(c) Skipped\_frames = 40

Tabla C.2: Resultados para la cámara lateral izquierda

#### ■ Cámara lateral derecha

Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
RT_FAR	18,3496	30,7334	0,5446	RT_FAR	16,0416	26,1609	0,5695
RT_CL	25,8609	45,948	0,4759	RT_CL	26,2923	46,743	0,4642
RT_ST	19,4907	33,3889	0,5558	RT_ST	14,9064	23,193	0,5909

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
RT_FAR	17,4476	29,081	0,5479
RT_CL	27,5081	48,6223	0,4586
RT_ST	24,8987	43,5069	0,4611

(c) Skipped\_frames = 40

Tabla C.3: Resultados para la cámara lateral izquierda

### C.0.0.2. Resultados Hipótesis de Estabilidad

#### ■ Cámara central



Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
CTR_FAR	16,7879	26,8832	0,5704	CTR_FAR	16,6896	26,1356	0,5736
CTR_CL	22,1082	38,392	0,4144	CTR_CL	22,0995	38,3346	0,4151
CTR_ST	26,4492	44,2727	0,3759	CTR_ST	25,9399	43,2183	0,3843

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
CTR_FAR	16,0762	25,9099	0,5803
CTR_CL	20,9503	35,1981	0,4402
CTR_ST	26,745	44,3793	0,367

(c) Skipped frames = 40

Tabla C.4: Resultados para la cámara central

#### ■ Cámara lateral izquierda

Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
LT_FAR	16,0525	24,5952	0,5576	LT_FAR	15,9181	24,3711	0,563
LT_CL	25,4278	41,7723	0,4049	LT_CL	22,8668	36,9348	0,4206
LT_ST	29,2313	48,9264	0,3384	LT_ST	30,4497	50,6162	0,3427

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
LT_FAR	14,8668	22,3883	0,5921
LT_CL	21,3879	34,4289	0,436
LT_ST	30,0222	50,2329	0,3377

(c) Skipped\_frames = 40

Tabla C.5: Resultados para la cámara lateral izquierda

#### ■ Cámara lateral derecha

Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
RT_FAR	18,0918	29,7714	0,5394	RT_FAR	16,0562	26,1798	0,5696
RT_CL	25,871	45,9609	0,476	RT_CL	26,4451	46,6433	0,4634
RT_ST	19,4976	33,3976	0,5559	RT_ST	14,9141	23,2033	0,591

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
RT_FAR	17,462	29,0936	0,548
RT_CL	27,5013	48,6183	0,4588
RT_ST	24,9127	43,525	0,4612

(c) Skipped\_frames = 40

Tabla C.6: Resultados para la cámara lateral derecha

### C.0.0.3. Sistema base

#### ■ Cámara central

Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
CTR_FAR	19,7166	32,0531	0,4534	CTR_FAR	20,8713	34,9696	0,412
CTR_CL	22,8675	38,2856	0,4396	CTR_CL	22,0457	36,1708	0,4212
CTR_ST	26,239	43,8893	0,3906	CTR_ST	27,2409	44,5815	0,3665

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
CTR_FAR	20,8924	35,4425	0,3919
CTR_CL	22,859	37,7561	0,4224
CTR_ST	25,9789	42,8666	0,3685

(c) Skipped frames = 40

Tabla C.7: Resultados para la cámara central

#### ■ Cámara lateral izquierda

Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
LT_FAR	17,2451	26,1027	0,5226	LT_FAR	16,7287	25,3912	0,5355
LT_CL	27,3235	44,8729	0,36	LT_CL	26,9323	44,1239	0,3662
LT_ST	29,5607	48,932	0,3491	LT_ST	30,4726	49,4833	0,3504

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
LT_FAR	16,0701	24,1836	0,5499
LT_CL	24,0296	38,8723	0,4064
LT_ST	31,6492	51,9827	0,3219

(c) Skipped\_frames = 40

Tabla C.8: Resultados para la cámara lateral izquierda

■ **Cámara lateral derecha**

Secuencia	AGE	pEPs	MS-SSIM	Nombre	AGE	pEPs	MS-SSIM
RT_FAR	11,678	17,7966	0,6602	RT_FAR	10,9208	16,2121	0,6928
RT_CL	29,6635	54,1996	0,3995	RT_CL	26,5024	58,2566	0,5219
RT_ST	17,9556	30,1801	0,5611	RT_ST	14,7414	22,9367	0,5929

(a) Resultados para skipped frames = 10

(b) Skipped frames = 20

Nombre	AGE	pEPs	MS-SSIM
RT_FAR	9,6086	12,6704	0,7532
RT_CL	0	0	0
RT_ST	0	0	0

(c) Skipped\_frames = 40

Tabla C.9: Resultados para la cámara lateral derecha

■ **Cámara central**

■ **Cámara lateral izquierda**

■ **Cámara lateral derecha**

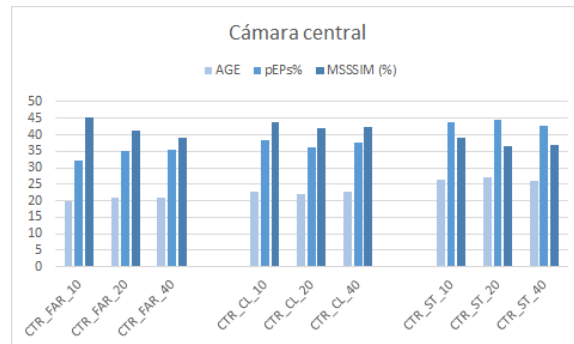


Figura C.1: Resultado de las métricas para la cámara central

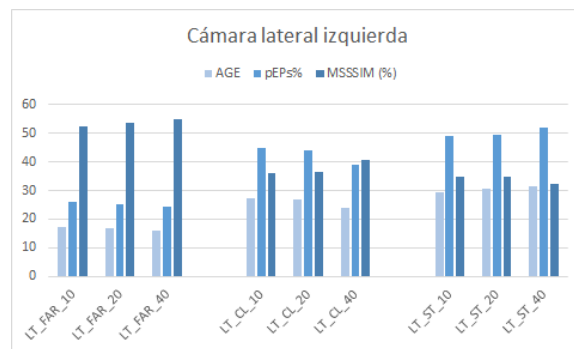


Figura C.2: Resultado de las métricas para la cámara lateral izquierda

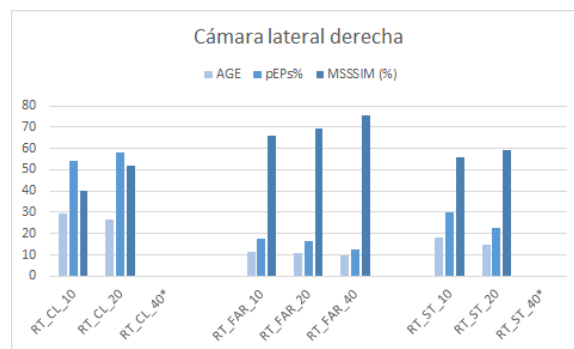


Figura C.3: Resultado de las métricas para la cámara lateral derecha

## Apéndice D

### Panorámicas creadas

Aquí se exponen, a mayor tamaño, las panorámicas que aparecen en la tabla X del capítulo Y.



Figura D.1: Panorámica generada a partir de la secuencia  $CT\_CL$ , mediante el sistema MP



Figura D.2: Panorámica generada a partir de la secuencia  $CT\_CL$ , mediante el sistema HE



Figura D.3: Panorámica generada a partir de la secuencia  $CT\_CL$ , mediante el sistema base



Figura D.4: Panorámica generada a partir de la secuencia  $LT\_FAR$ , mediante el sistema MP



Figura D.5: Panorámica generada a partir de la secuencia  $LT\_FAR$ , mediante el sistema HE



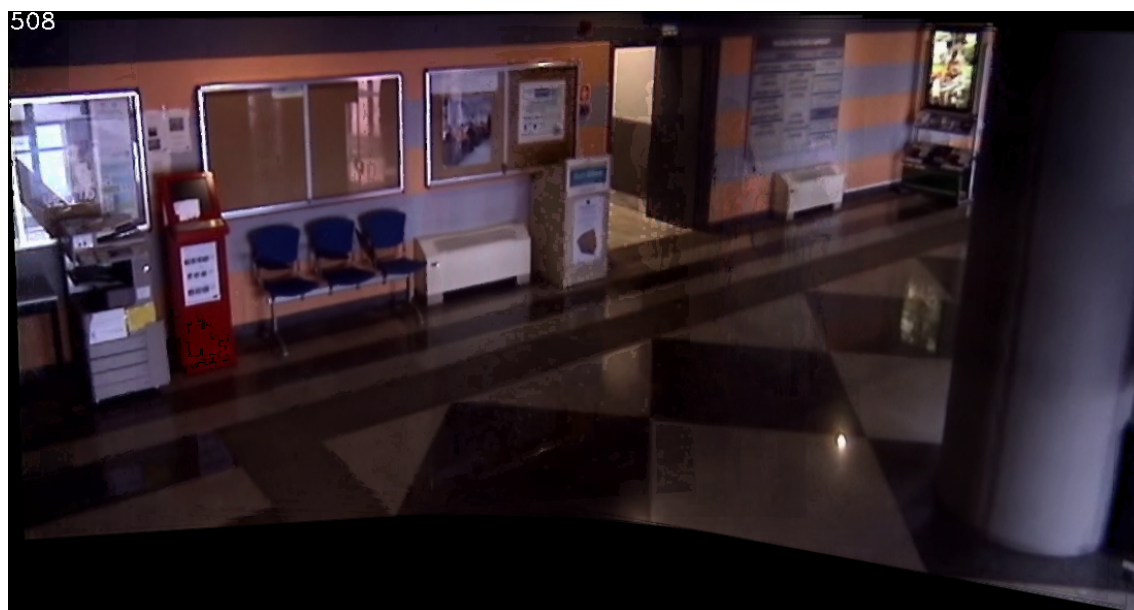


Figura D.6: Panorámica generada a partir de la secuencia  $LT\_FAR$ , mediante el sistema base

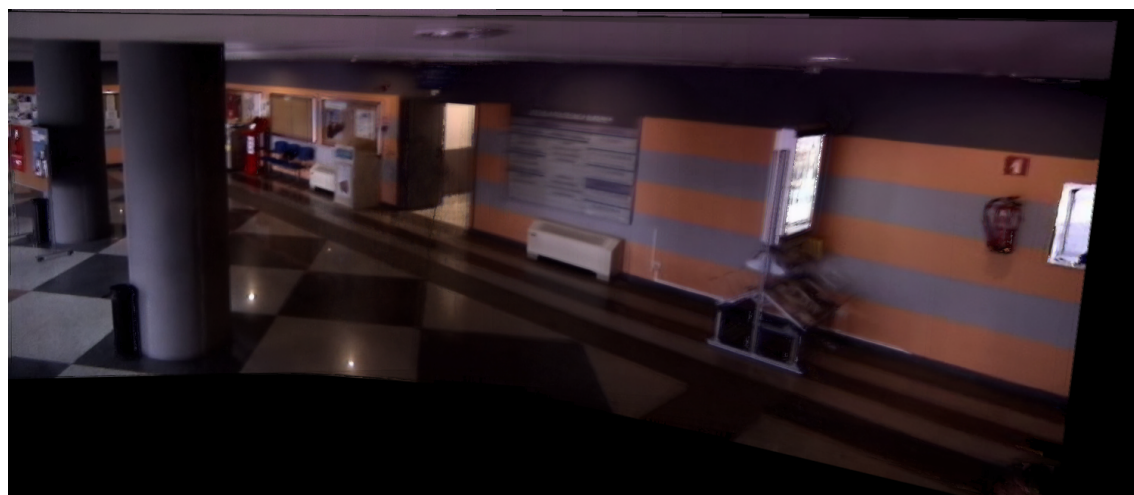


Figura D.7: Panorámica generada a partir de la secuencia  $RT\_CL$ , mediante el sistema MP



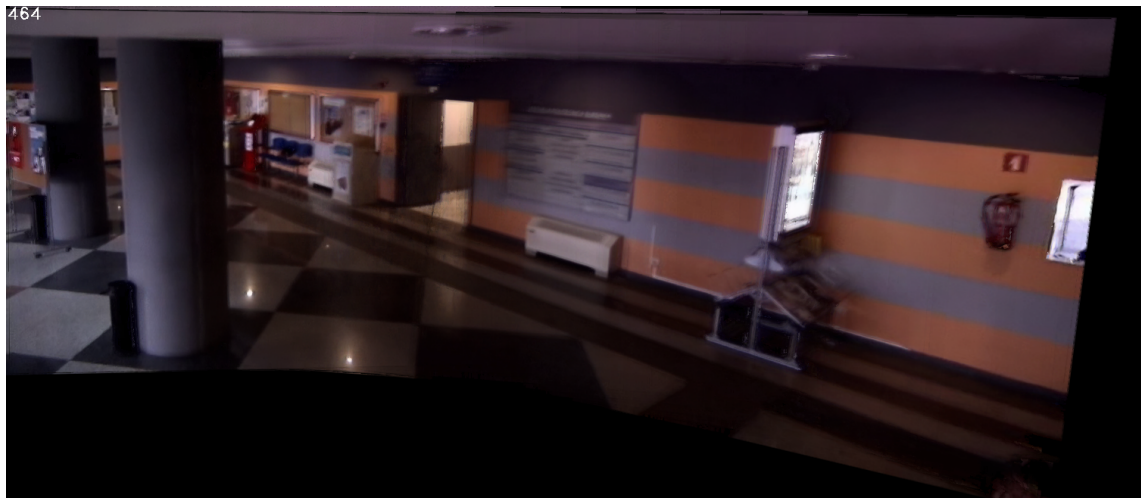


Figura D.8: Panorámica generada a partir de la secuencia  $RT\_CL$ , mediante el sistema HE

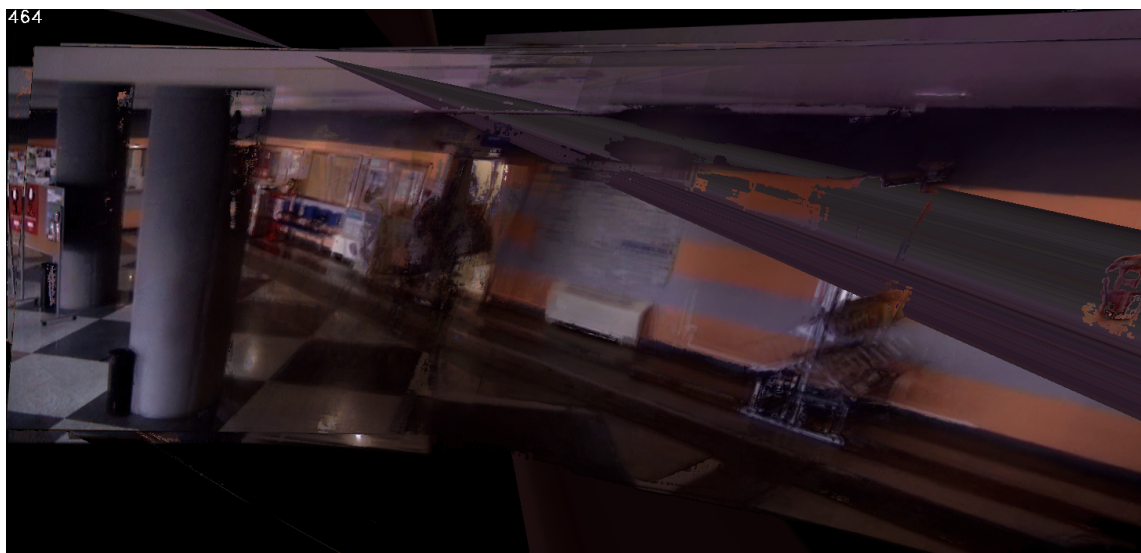


Figura D.9: Panorámica generada a partir de la secuencia  $RT\_CL$ , mediante el sistema base